Massachusetts Institute of Technology
Department of Economics
Working Paper Series

MARKETS VERSUS GOVERNMENTS: POLITICAL ECONOMY OF MECHANISMS

Daron Acemoglu,
Michael Golosov, Aleh Tsyvinski

Working Paper 06-09
April 24, 2006

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

Markets Versus Governments: Political Economy of Mechanisms*

Daron Acemoglu
MIT

Michael Golosov
MIT

Aleh Tsyvinski
Harvard

First Version: September 2005.
Current Version: April 2006.

Abstract

We study the optimal Mirrlees taxation problem in a dynamic economy with idiosyncratic (productivity or preference) shocks. In contrast to the standard approach, which implicitly assumes that the mechanism is operated by a benevolent planner with full commitment power, we assume that any centralized mechanism can only be operated by a self-interested ruler/government without commitment power, who can therefore misuse the resources and the information it collects. An important result of our analysis is that there will be truthful revelation along the equilibrium path (for all positive discount factors), which shows that truth-telling mechanisms can be used despite the commitment problems and the different interests of the government. Using this tool, we show that if the government is as patient as the agents, the best sustainable mechanism leads to an asymptotic allocation where the aggregate distortions arising from political economy disappear. In contrast, when the government is less patient than the citizens, there are positive aggregate distortions and positive aggregate capital taxes even asymptotically. Under some additional assumptions on preferences, these results generalize to the case when the government is benevolent but unable to commit to future tax policies. We conclude by providing a brief comparison of centralized mechanisms operated by self-interested rulers to anonymous markets.

Keywords: dynamic incentive problems, mechanism design, optimal taxation, political economy, revelation principle.

JEL Classification: H11, H21, E61, P16.

*We thank Manuel Amador, Marios Angeletos, Abhijit Banerjee, Tamer Basar, Timothy Besley, Olivier Blanchard, Ricardo Caballero, V. V. Chari, Mathias Dewatripont, Caroline Hoxby, Larry Jones, Patrick Kehoe, Narayana Kocherlakota, Jonathan Levin, David Levine, Guido Lorenzoni, Asuman Ozdaglar, Christopher Phelan, Andrew Postlewaite, Vasiliki Skreta, Robert Townsend, Pierre Yared, Muhamet Yildiz and seminar participants at Boston College, Boston University, Canadian Institute of Advanced Research, Carnegie Mellon, Cornell-Penn State Macro Conference, Federal Reserve Bank of Minneapolis, Federal Reserve Bank of Philadelphia Macroeconomics Conference, Federal Reserve Bank of Richmond, Georgetown, Harvard, MIT, Princeton, Rochester, Stanford, UCLA, Yale, University of Iowa, University of Kansas, and University of Pennsylvania for useful comments and to Oleg Itskhoki for excellent research assistance.

1 Introduction

The first-generation approach to public finance, best exemplified by models of Ramsey taxation, sought to determine the optimal policy of a benevolent government in a world with a given set of fiscal or regulatory instruments. The second-generation approach, pioneered by Mirrlees, has made major progress over this approach by explicitly modeling the choice of tax instruments that the government can use.[1] This literature has modeled the informational problems restricting the potential tax-transfer programs, and formulated the determination of optimal tax and transfer programs as one of mechanism design.[2]

Some of the theoretical limitations of the second-generation approach have long been apparent, however. First, in intertemporal settings, this approach assumes that the mechanism designer (government or planner) can *commit* to a dynamic mechanism, even though such commitment is ex post costly. Second, it is assumed that there is a body (a "benevolent government") that can operate the optimal tax-transfer program (mechanism), even though the lessons of the political economy literature are that governments or politicians do not simply maximize welfare, but have their own selfish objectives, such as reelection or personal enrichment.[3] This paper aims to contribute to a potential third-generation approach to public finance where both the informational constraints on tax instruments and the incentive problems associated with governments, politicians and bureaucrats are taken into account. For this purpose, we investigate how mechanisms work and should be designed in the presence of self-interested and time-inconsistent governments.[4]

Two questions motivate this analysis. First, we would like to understand whether *sustainable mechanisms*—i.e., optimal tax-transfer programs in the presence of these additional

[1]See Mirrlees (1971) for the seminal reference and Baron and Myerson (1982), Dasgupta, Hammond and Maskin (1979), Green and Laffont (1977), Harris and Townsend (1981), Myerson (1979), and Holmstrom and Myerson (1983) for some of the important papers in the early literature. Albanesi and Sleet (2005), Battaglini and Coate (2005), Golosov and Tsyvinski (2004), Golosov, Kocherlakota and Tsyvinski (2003), Kocherlakota (2005) and Werning (2002) consider applications of the Mirrlees framework to dynamic taxation.

[2]Given this formulation of the optimal tax-transfer program as a mechanism, we will use the terms "optimal tax-transfer program" and "mechanism" interchangeably.

[3]For general discussions of the implications of self-interested behavior of governments, petitions and bureaucrats, see, among others, Buchanan and Tullock (1962), North and Thomas (1973), North (1981), Olson (1982), North and Weingast (1989), and Dixit (2004). Austen-Smith and Banks (1999), Persson and Tabellini (2000) and Acemoglu (2005a) provide introductions to various aspects of the recent developments and the basic theory.

Another potential difficulty with centralized system is that they may involve excessive communication relative to trading systems. See Segal (2005) for a recent model developing this insight.

[4]One can think of an extended game in which there is a *fictional* disinterested mechanism designer, with the government as an additional player that has the authority to tax and regulate and the ability to observe all the communication between the fictitious mechanism designer and individual agents. Although this may be a useful modeling tool, it does not circumvent the substantive issues raised here: the party entrusted with taxes and transfers has neither the same interests as those of the citizens nor much commitment power. Naturally, the best mechanism we characterize can be represented as a solution to this fictional mechanism design problem.

constraints—look similar to "Mirrlees" mechanisms, which assume a benevolent planner and full commitment. If they do, then we can have more confidence in the mechanism design approach as a tool to analyze the practice of policy design as well as a normative benchmark.[5] Second, in the presence of self-interested and time-inconsistent government behavior, anonymous market allocations cannot always be replicated by centralized (sustainable) mechanisms. This opens the door to a theoretical analysis of the relative costs and benefits of centralized resource allocation methods versus those that preserve anonymity and limit government intervention, and we would like to take a first step in this direction.

To highlight the problems that arise when we depart from the benchmark of a benevolent planner with full commitment, it is useful to start with Roberts' (1984) example economy, where, similar to Mirrlees (1971), risk-averse individuals are subject to unobserved shocks affecting the marginal disutility of labor supply. But differently from the benchmark Mirrlees model, the economy is repeated T times, with individuals having perfectly persistent types. Under full commitment, a benevolent planner would choose the same allocation at every date, which coincides with the optimal solution of the static model. However, a benevolent government without full commitment cannot refrain from exploiting the information that it has collected at previous dates to achieve better risk sharing ex post. This turns the optimal taxation problem into a dynamic game between the government and the citizens. Roberts showed that as discounting disappears and $T \to \infty$, the unique sequential equilibrium of this game involves the highly inefficient outcome in which all types declare to be the worst type at all dates, supply the lowest level of labor and receive the lowest level of consumption. This example shows the potential inefficiencies that can arise once we depart from the unrealistic case of full commitment, even with benevolent governments.

Our benchmark economy incorporates the lack of commitment present in Roberts' (1984) paper, but also assumes that the government is self-interested and maximizes its own utility. This latter assumption brings our model closer to the political economy literature, where issues of conflict of interest and credibility of policy are central. It is also particularly useful for our purposes because it simplifies the structure of the dynamic game relative to the case with a benevolent time-inconsistent planner. At the end of the paper, we generalize some of our results to a situation where the government has an arbitrary degree of benevolence (thus nesting the fully-benevolent time-inconsistent government).

Our main departure from Roberts' (1984) framework is that instead of a finite-horizon

[5] In line with this objective, throughout the paper we look for the allocation that maximizes the ex ante utility of the citizens (agents) subject to the political economy and commitment constraints introduced by the self-interested nature of the government.

economy, we study an infinite-horizon economy, where individuals can use punishment strategies against the government. This enables us to construct a *sustainable mechanism*, defined as an equilibrium tax-transfer program that is both incentive compatible for the citizens and for the government (i.e., it satisfies a sustainability constraint for the government). The (best) sustainable mechanism gives a fraction of the output to the government in every period, and if the government deviates from this implicit agreement, citizens switch to supplying zero labor, implicitly punishing the government.[6] The infinite-horizon setup enables us to prove our first important result, that a version of the revelation principle, the *truthful revelation along the equilibrium path*, applies, irrespective of the discount factors of various parties. The fact that truthful revelation principle applies *only* along the equilibrium path is important, since it is actions off the equilibrium path that place restrictions on what types of mechanisms are allowed (these are encapsulated in the sustainability constraints).

The truthful revelation along the equilibrium path enables us to write the problem of finding the best sustainable mechanism as an infinite-dimensional maximization problem. We characterize the solution to this program by defining a *quasi-Mirrlees problem*, where the ex ante expected utility of the citizens is maximized subject to the standard incentive compatibility constraints and two additional resource constraints at every date; the first requires that the sum of total labor supply in the economy be no less than some amount L_t and the second that the sum of total consumption be no greater than some amount C_t. When the mechanism also optimizes over the sequences of C_t and L_t subject to the aggregate resource constraints, the quasi-Mirrlees problem is identical to the full-commitment dynamic Mirrlees problem. We show that the best sustainable mechanism is a solution to a quasi-Mirrlees problem, and distortions resulting from the self-interested behavior of the government only affect the parameters of this quasi-Mirrlees problem (i.e., L_t and C_t). This formulation therefore gives us a clean way of characterizing the differences between the best sustainable mechanism and the full-commitment Mirrlees mechanism in terms of the *aggregate distortions* caused by the former relative to the full-commitment Mirrlees allocation.

The other main results in the paper concern the characterization of these aggregate distortions. First, we show that at the initial date, there will always be further distortions in the

[6]Clearly, the punishment strategies that sustain this equilibrium are not "renegotiation proof". This is common with many other analyses of repeated games, so we do not view this as a special shortcoming of our approach. Moreover, in practice, citizens have many other recourses against governments that misbehave, including voting or throwing them out of office, and these actions will have the same impact as supplying zero labor in the punishment phase. We do not incorporate these possibilities to simplify the analysis in this paper. In Acemoglu, Golosov and Tsyvinski (2006), we study the equilibrium of an economy without any informational restrictions on taxes and transfers, where the citizens have the option of replacing the current government. This enables us to highlight the distortions arising from "pure political economy".

sustainable mechanism relative to the full-commitment Mirrlees mechanism. Intuitively, at the initial date, the sustainability constraint is never slack and implies that any increase in output has to be associated with increased government consumption. This increases the opportunity cost of increasing output, and leads to a reduction in labor supply and capital accumulation. Second, we provide tight conditions under which these further distortions disappear or persist over time. In particular, we prove that when the government is as patient as, or more patient than, the citizens, the sustainability constraint of the government eventually becomes slack. In the absence of a binding sustainability constraint, aggregate distortions disappear, and the marginal products of labor and capital asymptotically converge to the marginal rates of substitution for the agents; consequently, the limiting allocation can be implemented with a structure of taxes for labor similar to that in the full-commitment Mirrlees economy and zero aggregate capital taxes.[7] When the government is less patient than the citizens, the results are very different, however; aggregate distortions never disappear, and even in the long run, there are positive aggregate capital taxes. This last set of results is important, since it provides an exception to most existing models, which predict that long-run taxes on capital should be equal to zero (cfr. footnote 7).

The results for the case of a self-interested of government are derived under general assumptions on the utility functions of citizens. We also show that when individuals have instantaneous utility functions that are separable between consumption and leisure, similar results apply for any utility function of the government, in particular, in the case where the government is fully-benevolent but time-inconsistent.

Finally, since with a sustainable mechanism part of the output has to be given to the government, the anonymous market allocation, without the government, cannot always be achieved by a centralized mechanism. This raises the question of when centralized mechanisms are preferable to anonymous markets. We conclude the paper with a brief discussion of this issue; we show that anonymous markets with limited insurance and redistribution may be ex ante preferable to centralized mechanisms, and that this is more likely when there are worse institutional controls on government behavior and a lower discount factor of the government.

This paper is related to a number of different strands of research. These include both the original and the more recent applications of the mechanism design approach to the optimal taxation problem already mentioned in footnote 1. The major difference between our work and

[7] This result is therefore similar to that of zero limiting taxes on capital in the first-generation Ramsey-type models, e.g., Chamley (1986) or Judd (1985), but is derived here without any exogenous restriction on tax instruments (see Kocherlakota, 2005, for the zero capital tax result using the second-generation approach).

It is important to emphasize, however, that this limiting allocation can be decentralized in different ways, and some of those may involve positive taxes on individual capital holdings.

these papers is that they assume a benevolent government and full commitment. Secondly, our paper is related to the recently burgeoning political economy literature mentioned in footnote 3. What distinguishes our paper from this literature is the explicit modeling of the incentive problems on the side of the individuals as well as our focus on best sustainable mechanisms.[8] Our analysis is also related to work on optimal taxation with time-inconsistency, for example, Chari and Kehoe (1990, 1993), Phelan and Stacchetti (2001), and Sleet and Yeltekin (2004).

As well as Roberts (1984), most closely related to our research is the recent important paper by Bisin and Rampini (2005), who also consider the problem of mechanism design without commitment in a two-period setting.[9] Bisin and Rampini extend Roberts's analysis and show how the presence of anonymous markets acts as an additional constraint on the government, ameliorating the commitment problem. This lack of commitment is related to the lack of commitment by the self-interested government in our model. The most important distinction between the two approaches is that our model is infinite horizon. This enables us to construct sustainable mechanisms with the revelation principle holding along the equilibrium path, to analyze substantially more general environments, and to characterize the limiting behavior of distortions and taxes.

The rest of the paper is organized as follows. Section 2 describes the basic environment. Section 3 starts the analysis of sustainable mechanisms. In this section, we set up the problem of constructing sustainable mechanisms and prove a version of the revelation principle. Section 4 formulates the problem of characterizing the best sustainable mechanism as a solution to a quasi-Mirrlees program. Section 5 characterizes the best sustainable mechanism using a simple but restrictive case to illustrate the main ideas. Section 6 characterizes the best sustainable mechanism without any restrictions on the set of mechanisms and preferences. Section 7 extends our analysis to cover the case of fully-benevolent, but time-inconsistent governments. Section 8 briefly compares allocations with anonymous markets to those under sustainable mechanisms. Section 9 concludes, while the Appendices contain some technical material necessary for the analysis as well the proofs not provided in the text.

---

[8]In this context, Hart, Shleifer and Vishny (1997), Chari (2000) and Acemoglu, Kremer and Mian (2003) also contrast the incentive costs of governments and markets, but do not derive the costs of governments from the centralization of power and information in the process of operating a mechanism.

[9]See also the work by Freixas, Guesnerie and Tirole (1985) on the ratchet effect and recent work on general mechanisms without commitment, for example, Bester and Strausz (2001), Skreta (2004), and Miller (2005).

2 Demographics, Preferences and Technology

The model economy is infinite horizon in discrete time. It is populated by a continuum of citizens with measure normalized to 1 and a ruler.[10] The ruler/government can be thought of as a single agent or as a group of agents such as a bureaucracy, whose preferences can be consistently represented by a standard von Neuman-Morgenstern utility function.

Let $\Theta = \{\theta_0, \theta_1, ..., \theta_N\}$ be a finite ordered set of potential types, with the convention that θ_i corresponds to "higher skills" than θ_{i-1}, and in particular, θ_0 is the worst type.[11] Let Θ^T be the T-fold product of Θ, representing the set of sequences of length $T = 1, 2, ..., \infty$, with each element belonging to Θ. We think of each agent's lifetime type sequence θ^∞ as drawn from Θ^∞ according to some measure μ^∞. Let $\theta^{i,\infty}$ be the draw of individual i from Θ^∞. The t-th element of $\theta^{i,\infty}$, θ^i_t, is the skill level of this individual at time t. We use the standard notation $\theta^{i,t}$ to denote the history of this individual's skill levels up to and including time t, and make the standard measurability assumption that the individual only knows $\theta^{i,t}$ at time t. Since this will be a private information economy, no other agent in the economy will directly observe this history.[12] This structure imposes no restriction on the time-series properties of individual skills. Both iid draws from Θ in every period as well as arbitrary temporal dependence are allowed. For concreteness, one may wish to think that θ^i_t follows a Markov process. We assume that each individual's lifetime type sequence is drawn from Θ^∞ according to the same measure μ^∞ and independently from the draws of all other individuals, so that there is no aggregate uncertainty in the skills distribution. In addition, to simplify the notation, we also assume (without loss of generality) that within each period, there is an aggregate invariant distribution of types denoted by G.

The instantaneous utility function of individual i at time t is given by

$$u\left(c^i_t, l^i_t \mid \theta^i_t\right) \tag{1}$$

where $c^i_t \geq 0$ is the consumption of this individual and $l^i_t \geq 0$ is her labor supply. This

[10] The continuum assumption implies that whenever we think of deviations, these should be by an individual with positive measure ε, and then we should take the limit as $\varepsilon \to 0$. Moreover, equilibrium statements should be read as "almost-everywhere". These technical details do not matter except in the proof of Theorem 5.

[11] Finiteness of Θ is adopted for simplicity and without loss of any economic insight. The more general case where Θ is a compact interval of $\mathbb{R}_+$ introduces a number of additional technical details, not central for our analysis.

[12] This means that there exists a set of nested information sets (sub-sigma fields) representing each individual's information sets, so that the individual only knows the information contained in $\mathcal{F}^i_t$ at time t. In particular, let the triple $(\Theta^\infty, \mathcal{F}, \mu^\infty)$ be a probability space and $\left\{\mathcal{F}^i_t : t \in \mathbb{Z}_+\right\}$ be a filtration, i.e., a collection of sub-sigma fields of $\mathcal{F}$, such that $\mathcal{F}^i_t \subseteq \mathcal{F}^i_{t'}$ for all $t' > t$. Let Θ^t be the set Θ^∞ truncated at t. Then $\theta^{i,t} \in \Theta^t$ and all decisions taken at time t by individual i must be $\mathcal{F}^i_t$-measurable.

formulation is general enough to nest both preference shocks and productivity shocks.[13] We assume that labor supply of an individual with skill θ comes from a compact set, i.e., $l_t^i \in [0, \bar{l}(\theta)]$.

Assumption 1 *(utility function) For all $\theta \in \Theta$, $u(c, l \mid \theta) : \mathbb{R}_+ \times [0, \bar{l}(\theta)] \to \mathbb{R}$ is twice continuously differentiable and jointly concave in c and l, and is non-decreasing in c and non-increasing in l.*

Assumption 2 *(single crossing) Let the partial derivatives of u be denoted by u_c and u_l. Then $u_c(c, l \mid \theta) / |u_l(c, l \mid \theta)|$ is increasing in θ for all c and l and all $\theta \in \Theta$.*

Assumption 3 *(worst type and full support) We have $\bar{l}(\theta_0) = 0$ and $\bar{l}(\theta) = \bar{l} < \infty$ for all $\theta \in \Theta$ and $\theta \neq \theta_0$. Moreover, μ^∞ has full support in the sense that $\theta_t^i = \theta_0$ has positive probability after any history.*

The first two assumptions are standard. Assumption 3 states that for the worst type, θ_0, supplying positive labor is impossible. This suggests that we can think of the worst type as "disabled"—unable to supply any labor at that date. It also requires μ^∞ to have full support in the sense that any individual can become disabled at any point. This assumption will simplify the analysis of sustainable mechanisms by making it possible to have off-the-equilibrium path actions where all types supply zero labor. As described in Remark 1, this assumption can be replaced by an alternative one, described below, which imposes "freedom of labor supply" directly. The advantage of Assumption 3 is that it leads to the freedom of labor supply as an equilibrium outcome.

Each individual maximizes the discounted sum of their utility with discount factor $\beta \in (0, 1)$, so their objective function at time t is

$$\mathbb{E}\left[\sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}^i, l_{t+s}^i \mid \theta_{t+s}^i\right) \middle| \mathcal{F}_t^i\right] = \mathbb{E}\left[\sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}^i, l_{t+s}^i \mid \theta_{t+s}^i\right) \middle| \theta^{i,t}, h^t\right]$$

where $\mathbb{E}\left[\cdot | \mathcal{F}_t^i\right]$ or $\mathbb{E}\left[\cdot | \theta^{i,t}, h^t\right]$ denote the expectations operator conditional on having observed the history $\theta^{i,t}$ in addition to any public information up to time t, captured by h^t (discussed further below).

The production side of the economy is described by the aggregate production function

$$Y = F(K, L) \tag{2}$$

where K is capital and L is labor. We assume:

[13] In particular, productivity shocks would correspond to the case where $u\left(c_t^i, l_t^i \mid \theta_t^i\right) = u\left(c_t^i, l_t^i / \theta_t^i\right)$.

Assumption 4 *(production structure) F is strictly increasing and continuously differentiable in both of its arguments, with derivatives denoted by F_K and F_L, exhibits constant returns to scale and satisfies the Inada condition, $\lim_{K\to\infty} F_K(K,L) = 0$ for all $L \in \mathbb{R}_+$. Moreover, capital fully depreciates after use, and $F(K,0) = 0$.*

Both the full depreciation assumption and the assumption that labor is essential for production are adopted to simplify the notation. The Inada condition, together with the fact that the maximum amount of labor in the economy is bounded, implies that there is a maximum amount of output that can be produced $\bar{Y} \in (0,\infty)$ given by $\bar{Y} = F\left(\bar{Y}, \bar{L}\right)$, where $\bar{L}$ is the maximum amount of total labor.

In addition, the ruler's (government's) utility at time t is given by

$$\mathbb{E}_t\left[\sum_{s=0}^{\infty} \delta^s v\left(x_{t+s}\right)\right],$$

where x denotes government consumption, $v : \mathbb{R}_+ \to \mathbb{R}$ is the ruler's instantaneous utility function, and $\mathbb{E}_t$ refers to the expectations operator conditional on public information at time t. We assume that:

Assumption 5 *(ruler utility) v is twice continuously differentiable, concave, and satisfies $v'(x) > 0$ for all $x \in \mathbb{R}_+$ and $v(0) = 0$. Moreover $\delta \in (0,1)$.*

Notice also that the ruler's discount factor, δ, is potentially different from that of the citizens, β.

3 Sustainable Mechanisms

Since the government both lacks commitment power and has the ability to expropriate output for its own consumption, the interaction between the citizens and the government is a game (e.g., Roberts, 1984, Chari and Kehoe, 1990, 1993). Our purpose throughout is to characterize the equilibrium of this game between the government and the citizens, corresponding to the *best sustainable mechanism*, meaning the sustainable mechanism that maximizes the ex ante utility of citizens.[14]

[14] Since we are dealing with a dynamic game, our focus on the best sustainable mechanism is essentially a selection among the many equilibria. Alternatively, one can think of the "social plan" as being designed by the citizens to maximize their utility subject to the constraints placed by the self-interested behavior of the government (see, in particular, the last paragraph of the Concluding Remarks, and also Acemoglu, 2005). In addition, throughout the paper we focus on perfect Bayesian equilibria (see Definition 1) and do not impose renegotiation-proofness (focusing on the renegotiation-proof subset of equilibria is both technically and conceptually difficult and also would put more constraints on the "best sustainable mechanism").

This section sets up the details of the game between the government and the citizens, and proves Proposition 1, which provides an infinite-dimensional maximization problem for the characterization of the best sustainable mechanism. This proposition is then used to determine the form of the best sustainable mechanism and the resulting distortions. This section will also establish a general result on truthful revelation along the equilibrium path (Theorem 1, also Theorem 5 below), which is used here to prove Proposition 1, but is of independent interest for infinite-horizon games and mechanism design problems without commitment, and shows that these problems can be analyzed without giving up the revelation principle.

3.1 The Game Form Between Government and Citizens

There is a number of alternative ways of specifying the game form between the government and the citizens, with identical results. Our choice here is motivated to maximize similarity with Roberts' (1984) model.

We define a *submechanism* (or t-mechanism) as a subcomponent of the overall mechanism between the government and the individuals. A submechanism specifies what happens at a given date. In particular, let Z_t be a general message space for time t, with a generic element z_t.[15] This message space may include messages about current type of the individual, $\hat{\theta}_t^i \in \Theta$, and past types $\hat{\theta}^{i,t-1} \in \Theta^{t-1}$ (even though the individual may have made some different reports about his or her types in the past), and might also include other messages.

Let $Z^t \equiv \prod_{s=0}^{t} Z_s$ and z^t denote a generic element of Z^t. In addition, let $h_t \in H_t$ be some publicly observable aggregate variable, and $h^t \in H^t \equiv \prod_{s=0}^{t} H_s$ denotes the history of this variable up to time t. Here h_t could be payoff relevant or simply a "sunspot" variable used for randomization, and is *not* affected by the actions of the agents. These histories are introduced to allow for randomizations.

A submechanism consists of two mappings, i.e., $M_t \equiv \left(\tilde{c}_t, \tilde{l}_t\right)$ such that $\tilde{c}_t : Z^t \times H^t \to \mathbb{R}_+$ assigns consumption levels for each complete history of messages and public histories, and $\tilde{l}_t : Z^t \times H^t \to \left[0, \bar{l}\right]$ assigns corresponding labor supply levels.[16] Given Assumption 3, any

[15]More formally, θ^t, $\hat{\theta}^t$ and z_t have to be $\mathcal{F}_t$-measurable as defined in footnote 12.

[16]The mechanisms we describe here allow for general message spaces, but impose two restrictions. First, they are non-stochastic. This is only to simplify notation in the text, and everywhere we can replace the mappings in the text with $\tilde{c}_t : Z^t \times H^t \to \Delta(\mathbb{R}_+)$ and $\tilde{l}_t : Z^t \times H^t \to \Delta\left(\left[0, \bar{l}\right]\right)$, where $\Delta(J)$ denotes the set of probability measures over J. In the Appendix, we consider potentially stochastic mechanisms to convexify the constraint set. Second, a more general mechanism would be a mapping from the message histories of all agents, not just the individual's history. Since there is a continuum of agents that do not share any information, this latter restriction is without loss of generality here (except that off the equilibrium path, some submechanisms would violate the resource constraint, though this is not important for our equilibrium analysis). Notice also that while the submechanism restricts each individual's allocations to be a function of only his own history of reports, as

submechanism must allow for some messages which will lead to $l = 0$. We denote the set of submechanisms that satisfy this restriction and also the relevant resource constraints (which will be specified below) by $\mathcal{M}_t$. Throughout, we condition strategies on public history h^t whenever such conditioning is useful, and suppress this dependence when the context makes the conditioning clear.

The typical assumption in models with no commitment is that the mechanism designer can commit to a submechanism at a given date, but cannot commit to what mechanisms will be offered in the future. In our context, there is an additional type of deviation for the government whereby it can use its power to extract resources from the society even within the same period. The interaction between the government and the individuals is modeled with the following game form at each date:

1. At the beginning of period t, the government offers a submechanism $\tilde{M}_t \in \mathcal{M}_t$.

2. Individuals send a message $z_t \in Z_t$, and $h_t \in H_t$ is realized; z_t and h_t together with $h^{t-1} \in H^{t-1}$ and $z^{t-1} \in Z^{t-1}$ determine labor supplies according to the submechanism $\tilde{M}_t$.

3. Production takes place according to the labor supplies of the individuals, with $Y_t\left(h^t\right) = F\left(K_t, L_t\left(h^t\right)\right)$, where K_t is the capital stock inherited from the previous period, and $L_t\left(h^t\right) = \int_0^1 \tilde{l}_t\left(z^{i,t}, h^t\right) di$, where $i \in [0,1]$ indexes individuals and $z^{i,t} \in Z^t$ denotes the history of reports by individual i.

4. The government decides whether to deviate from the submechanism $\tilde{M}_t$, denoted by $\xi_t\left(h^t\right) \in \{0,1\}$. If $\xi_t\left(h^t\right) = 0$, production is distributed among agents according to the pre-specified submechanism $\tilde{M}_t \in \mathcal{M}_t$, the government chooses $\tilde{x}_t\left(h^t\right) \leq F\left(K_t, L_t\left(h^t\right)\right)$, and next period's capital stock is determined as $\tilde{K}_{t+1}\left(h^t\right) = F\left(K_t, L_t\left(h^t\right)\right) - \tilde{x}_t\left(h^t\right) - \int_0^1 \tilde{c}_t\left(z^{i,t}, h^t\right) di$. If $\xi_t\left(h^t\right) = 1$, the government chooses $\tilde{x}'_t\left(h^t\right) \leq F\left(K_t, L_t\left(h^t\right)\right)$, and a new consumption function $\tilde{c}'_t : Z^t \times H^t \to \mathbb{R}_+$, and next period's capital stock is: $\tilde{K}'_{t+1}\left(h^t\right) = F\left(K_t, L_t\left(h^t\right)\right) - \tilde{x}'_t\left(h^t\right) - \int_0^1 \tilde{c}'_t\left(z^{i,t}, h^t\right) di$ for all $h^t \in H^t$.[17]

---

it will become clear below, the government's strategies allow submechanisms to be functions of the reports of *all* agents in the past.

Finally, we could define a submechanism as a mapping $M_t[K_t]$ conditional on the capital stock of the economy at that date to emphasize that what can be achieved will be a function of the capital stock. We suppress this dependence to simplify notation.

[17]More generally, we can allow the government to capture a fraction $\eta \leq 1$ of the total output of the economy when $\xi = 1$, where the level of η could be related to the institutional controls on government or politician behavior. In this case, the constraint on the government following a deviation would be $\tilde{K}'_{t+1}\left(h^t\right) = \eta F\left(K_t, L_t\left(h^t\right)\right) - \tilde{x}'_t\left(h^t\right) - \int_0^1 \tilde{c}'_t\left(z^{i,t}, h^t\right) di$, with the remaining $1 - \eta$ fraction of the output

This game form emphasizes that the only difference between the standard models with no commitment and our setup is that the government, in the last stage, can also decide to expropriate the output produced in the economy. Notice that at this stage, labor supply decisions have already been made according to the pre-specified submechanism $\tilde{M}_t$. However, consumption allocations cannot be made according to $\tilde{M}_t$, since the government is expropriating some of the output for itself. Consequently, we also let the government choose a new consumption allocation function, $\tilde{c}_t' : Z^t \times H^t \to \mathbb{R}_+$ at this point.

Let $M = \{M_t\}_{t=0}^{\infty}$ with $M_t \in \mathcal{M}_t$ be a mechanism, with the set of mechanisms denoted by $\mathcal{M}$. Let $x = \{x_t(h^t)\}_{t=0}^{\infty}$ be the (potentially stochastic) sequence of government consumption levels. We define a *social plan* as (M, x), which is an implicitly-agreed sequence of submechanisms and consumption levels for the government.

We represent the action of the government at time t conditional on history of publicly observable variables, h^t, by $\rho_t = \left(\tilde{M}_t, \xi_t(h^t), \tilde{x}_t(h^t), \tilde{x}_t'(h^t), \tilde{c}_t'\right)$. The first element of ρ_t is the submechanism that the government offers at stage 1 of time t, and the second is the government's expropriation decision. The third element of ρ_t is what the government consumes itself if $\xi_t = 0$. Since $\tilde{M}_t$ specifies both total production and total consumption by the citizens, given $\tilde{x}_t$ the capital stock for next period, $\tilde{K}_{t+1}$, is determined as a residual from the resource constraint and is not specified as part of the action profile of the government.[18] The fourth element, $\tilde{x}_t$, is the government consumption level when $\xi_t = 1$. Finally, the fifth element is the function $\tilde{c}_t'$ that the government chooses after deviating from the original submechanism, with $\mathcal{C}_t$ denoting the set of all such functions. Once again the capital stock for the following period, $\tilde{K}_{t+1}'$, is determined as a residual from the resource constraint. Government consumption levels must satisfy: $\tilde{x}_t \leq F(K_t, L_t)$ and $\tilde{x}_t' \leq F(K_t, L_t)$, but to simplify notation we write $\tilde{x}_t$, $\tilde{x}_t' \in \mathbb{R}_+$. Let $\mathcal{R}_t$ be the set of ρ_t's and $\rho^t \in \mathcal{R}^t$ denote the history of ρ_t's up to and including time t, and assume that this is publicly observable.[19]

For the citizens, define $\alpha_t^i(\theta^t \mid z^{t-1}, \rho^{t-1}, h^{t-1})$ as the action of individual i at time t when her type history is θ^t, her history of messages so far is z^{t-1} and the publicly observed histories of government actions and aggregate variables up to time $t-1$ are ρ^{t-1} and h^{t-1}. The action

getting destroyed. This generalization has no effect on our results and for now, we set $\eta = 1$ to simplify notation. We return to issues of institutional limits on government expropriation below.

[18] Since we are characterizing a (sustainable) mechanism, the ownership of the capital stock $\tilde{K}_{t+1}$ is not specified. Instead, this is simply the amount of resources used in production in the following period, and the government decides how this production will be distributed.

[19] In fact, ρ^t includes the action $\tilde{x}_t'$ and the function $\tilde{c}_t'$, which are not observed when $\xi_t = 0$. Thus, more appropriately, only a subset of ρ^t should be observed publicly. This slight abuse of notation is without any consequence for the analysis.

α_t^i specifies a message $z_t \in Z_t$, so:

$$\alpha_t^i : Z^{t-1} \times \mathcal{R}^{t-1} \times \Theta^t \times H^{t-1} \to Z_t.$$

We write $z^t\left(\alpha_t\left(\theta^t, h^{t-1}\right)\right)$ to denote the message resulting from strategy α_t for an agent of type θ^t and given public histories h^{t-1}. A strategy is *truth telling* if it satisfies

$$\alpha^*\left(\theta^t \mid z^{t-1}, \rho^{t-1}, h^{t-1}\right) = z_t\left[\theta^t\right] \text{ for all } \theta^t \in \Theta^t,\ z^{t-1} \in Z^{t-1},\ \rho^{t-1} \in \mathcal{R}^{t-1} \text{ and } h^{t-1} \in H^{t-1}, \tag{3}$$

where the notation $z_t\left[\theta^t\right]$ means that the individual is sending a message that fully reveals her true type.[20] To economize on notation, we represent the truth-telling strategy by $\alpha_t^i\left(\theta_t \mid z^{t-1}\left[\theta^{t-1}\right], \rho^{t-1}, h^{t-1}\right) = \alpha^*$. Notice that this strategy only imposes truth-telling following truthful reports in the past (since instead of an arbitrary history of messages z^{t-1}, we have conditioned on $z^{t-1}\left[\theta^{t-1}\right]$). In addition, let us define the null strategy

$$\alpha^{\emptyset}\left(\theta_t \mid z^{t-1}, \rho^{t-1}, h^{t-1}\right) = z_t^{\emptyset} \text{ for all } \theta^t \in \Theta^t,\ z^{t-1} \in Z^{t-1},\ \rho^{t-1} \in \mathcal{R}^{t-1} \text{ and } h^{t-1} \in H^{t-1},$$

where $z_t^{\emptyset}$ stands for a message signifying that the individual is disabled (i.e., $\theta_t^i = \theta_0$). Such a message must always be allowed in any submechanism that is an element of $\mathcal{M}_t$ because of Assumption 3.[21] Therefore, the individual can always choose to supply zero labor, or in other words, any feasible mechanism (submechanism) must allow for "freedom of labor supply". We will use the notation $\alpha_t^i\left(\theta_t \mid z^{t-1}, \rho^{t-1}, h^{t-1}\right) = \alpha^{\emptyset}$ to denote that the individual is playing the null strategy. Finally, we denote the *strategy profile* of all the individuals in society by $\underline{\alpha}$, with A denoting the set of all such strategy profiles.

Let $\underline{z}_t \in \mathcal{Z}_t$ be a profile of reports at time t.[22] As usual, we define $\mathcal{Z}^t = \prod_{s=0}^{t} \mathcal{Z}_s$. The government's strategy at time t is therefore

$$\Gamma_t : \mathcal{R}^{t-1} \times \mathcal{Z}^{t-1} \times H^{t-1} \to \mathcal{R},$$

i.e., it determines $\tilde{M}_t \in \mathcal{M}_t$, $\xi_t \in \{0,1\}$, $\tilde{x}_t \in [0, F(K_t, L_t)]$, $\tilde{x}_t' \in [0, F(K_t, L_t)]$ and $\tilde{c}_t \in \mathcal{C}_t$ as a function of the government's own past actions and the entire history of reports by citizens. We denote the strategy of the government by Γ and the set of government strategies by $\mathcal{G}$.

[20] Whenever we write "for all $h^{t-1} \in H^{t-1}$," this should be understood as "almost surely," since some histories may have zero probability.

[21] Since an individual with $\theta_t^i = \theta_0$ cannot supply any labor, he must always send the message $z_t^{\emptyset}$ (or an equivalent message).

[22] More formally, $\underline{z}_t$ assigns a report to each individual, thus it is a function of the form $\underline{z} : [0,1] \to Z_t$, where $i \in [0,1]$ denotes individual i, and $\mathcal{Z}_t$ is the set of all such functions.

Definition 1 *A (*Perfect Bayesian*)* equilibrium *in the game between the government and the citizens is given by strategy profiles* $\hat{\Gamma}$ *and* $\underline{\alpha}$ *that are* sequentially rational, *i.e., best responses to each other in all information sets given beliefs, and whenever possible, beliefs are derived from Bayesian updating given the strategy profiles.*[23] *We write the requirement that these strategy profiles are best responses to each other as* $\hat{\Gamma} \succeq_{\underline{\alpha}} \Gamma$ *for all* $\Gamma \in \mathcal{G}$ *and* $\underline{\alpha} \succeq_{\hat{\Gamma}} \underline{\alpha}$ *for all* $\underline{\alpha} \in \mathbf{A}$.

Let us define $\Gamma_{M,x} = \left[\left\{\tilde{M}_t, \xi_t\left(h^t\right), \tilde{x}_t\left(h^t\right), \tilde{x}_t'\left(h^t\right), \tilde{c}_t'\right\}_{t=0}^{\infty}\right]$ as the (potentially stochastic) action profile of the government induced by strategy Γ given a social plan (M, x).

Definition 2 M *is a sustainable mechanism if there exists* $x = \left\{x_t\left(h^t\right)\right\}_{t=0}^{\infty}$, *a strategy profile* $\underline{\alpha}$ *for the citizens and a strategy profile* $\Gamma_{M,x} \in \mathcal{G}$ *for the government, which constitute an equilibrium and induce an action profile* $\left[\left\{\tilde{M}_t, \xi_t\left(h^t\right), \tilde{x}_t\left(h^t\right), \tilde{x}_t'\left(h^t\right), \tilde{c}_t'\right\}_{t=0}^{\infty}\right]$ *for the government such that* $\tilde{M}_t = M_t$, $\xi_t\left(h^t\right) = 0$, *and* $\tilde{x}_t\left(h^t\right) = x_t\left(h^t\right)$ *for all* $h^t \in H^t$, *and satisfies* $\Gamma_{M,x} \succeq_{\underline{\alpha}} \Gamma$ *for all* $\Gamma \in \mathcal{G}$. *In this case, we say that equilibrium strategy profiles* $\Gamma_{M,x}$ *and* $\underline{\alpha}$ *support the sustainable mechanism* M.

In essence, this implies that the government does not wish to deviate from the social plan (M, x) given the strategy profile, $\underline{\alpha}$, of the citizens. The notation $\hat{\Gamma} \succeq_{\underline{\alpha}} \Gamma$ makes this explicit, stating that given the strategy profile, $\underline{\alpha}$, of the citizens, the government weakly prefers its strategy profile to any other strategy profile based on the same implicit agreement.

3.2 Truthful Revelation Along the Equilibrium Path

The revelation principle is a powerful tool for the analysis of mechanism design and implementation problems (see, e.g., MasCollel, Winston and Green, 1995). Since, in this environment, the government, who operates the mechanism, cannot commit and has different interests than those of the agents, the simplest version of the revelation principle does not hold; there will exist situations in which individuals will prefer not to report their true type (e.g., Roberts, 1984, Freixas, Guesnerie and Tirole, 1985, or Bisin and Rampini, 2005).[24] The key result of this section will be that along the equilibrium path, a version of the revelation principle will hold (without introducing a fictional mechanism designer and for all positive discount factors).

Let us first consider the problem of finding the best allocation for individuals. As we will see below, as long as the set of sustainable mechanisms (i.e., the constraint set, (5)-(7)) is

[23] We do not introduce explicit notation to describe beliefs, since these do not play any role in any of the analysis or the proofs.

[24] As noted in the Introduction, this statement refers to the case in which messages are sent to the government. It is possible to construct alternative environments with fictional mechanism designers with full commitment power, so that the revelation principle holds.

nonempty, this is equivalent to choosing the best sustainable mechanism, given by the following program:

$$\max \mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t u\left(\tilde{c}_t\left(z^t\left[\alpha_t\left(\theta^t,h^{t-1}\right),h^t\right]\right),\tilde{l}_t\left(z^t\left[\alpha_t\left(\theta^t,h^{t-1}\right),h^t\right]\right)\mid\theta_t^i\right)\right] \tag{4}$$

subject to the resource constraint,

$$\begin{aligned} K_{t+1}\left(h^t\right) &= F\left(K_t\left(h^{t-1}\right),\int\tilde{l}_t\left(z^t\left[\alpha_t\left(\theta^t,h^{t-1}\right)\right],h^t\right)dG^t\left(\theta^t\right)\right) \\ &\quad -\int\tilde{c}_t\left(z^t\left[\alpha_t\left(\theta^t,h^{t-1}\right)\right],h^t\right)dG^t\left(\theta^t\right)-\tilde{x}_t\left(h^t\right), \end{aligned} \tag{5}$$

a set of incentive compatibility constraints for individuals,

$$\underline{\alpha} \text{ is a best response to } \Gamma_{M,x}, \tag{6}$$

and the "sustainability" constraint of the government:

$$\mathbb{E}\left[\sum_{s=0}^{\infty}\delta^s v\left(\tilde{x}_{t+s}\left(h^{t+s}\right)\right)\middle| h^t\right]\geq \max_{\tilde{x}'_t,\tilde{K}'_{t+1},\tilde{c}'_t}\mathbb{E}\left[\left\{v\left(\tilde{x}'_t\right)+\delta v_t^c\left(\tilde{K}'_{t+1},\tilde{c}'_t\mid\tilde{M}^t\right)\right\}\mid h^t\right], \tag{7}$$

for all $t\geq 0$ and all $h^t\in H^t$. Note that current labor supply decisions are conditioned on the realization of the public history up to time t, h^t, while the capital stock, K_t, inherited from the previous period, is conditioned on h^{t-1}.

The last constraint, (7), encompasses all the possible deviations by the government at date t: the left-hand side is what the government will receive from date t onwards by sticking with the implicitly-agreed consumption schedule for itself. The right-hand side is the maximum it can receive by deviating. The potential deviations include a deviation at the last stage of the subgame at time t to expropriation, $\xi_t = 1$, together with a new consumption schedule for individuals, $\tilde{c}'_t$; or $\xi_t = 0$ and a choice of $\tilde{x}_t\left(h^t\right)$ different from $x_t\left(h^t\right)$; or the offer of a new submechanism at time $t+1$ (encapsulated into the continuation value v_t^c). In the case where $\xi_t = 1$, the government chooses $\tilde{x}'_t$, $\tilde{K}'_{t+1}$ and $\tilde{c}'_t$ to maximize its deviation value, which is given by current utility, $v\left(\tilde{x}_t\right)$, and continuation value, written as $v_t^c\left(\tilde{K}'_{t+1},\tilde{c}'_t\mid\tilde{M}^t\right)$, to emphasize that this continuation value depends on the entire history of submechanisms (thus information) up to time t, $\tilde{M}^t$, and on the capital stock from then on, $\tilde{K}'_{t+1}$, as well as potentially on $\tilde{c}'_t$. If this constraint, (7), were not satisfied, it is either because the government prefers $\xi_t = 0$ and some sequence of submechanisms or consumption levels different from (M,x), or because the government prefers $\xi_t = 1$. In the former case, we can always change (M,x) to ensure that (7) is satisfied. The latter, i.e., $\xi_t = 1$, cannot be part of the best equilibrium allocation from the

viewpoint of the citizens, since it involves government expropriation. Consequently, as long as the constraint set given by (5)-(7) is nonempty, the best allocation must satisfy (7) and is thus a solution to the program of maximizing (4) subject to (5)-(7). Finally, this constraint set is indeed nonempty, since the trivial allocation with zero production and zero consumption for all parties is in the set.[25]

Let us also introduce the notation $\underline{\alpha}= (\alpha \mid \alpha')$ to denote a strategy profile where all individuals play α along the equilibrium path and α' off the equilibrium path. We then have:

Lemma 1 *Suppose Assumptions 1-5 hold. Then in any sustainable mechanism,*

$$\mathbb{E}\left[\sum_{s=0}^{\infty}\delta^{s}v\left(x_{t+s}\left(h^{t+s}\right)\right)\middle| h^{t}\right]\geq v\left(F\left(K_{t}\left(h^{t-1}\right),L_{t}\left(h^{t}\right)\right)\right)\text{ for all }t\text{ and }h^{t}\in H^{t},\tag{8}$$

is necessary. Moreover, in the best sustainable mechanism, $v_t^c\left(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t\right) = 0$ for all $\tilde{M}^t \in \mathcal{M}^t$, $\tilde{K}'_{t+1} \in \mathbb{R}_+$ and $\tilde{c}'_t \in \mathcal{C}_t$, and the sustainability constraint (7) is equivalent to (8).

Proof. See Appendix B. ■

This lemma uses the fact that irrespective of the history of submechanisms and the amount of capital stock left for future production, there is an equilibrium continuation play that gives the government zero utility from that point onwards (which is analogous to the results in repeated games where the most severe punishments against deviations are optimal, e.g., Abreu, 1988). This continuation play is used as the threat against government deviation from the implicitly-agreed social plan. The implication is that, along the best sustainable mechanism, the best deviation for the government involves $\xi_t = 1$ and $\tilde{x}'_t = F(K_t, L_t)$. This enables us to simplify the sustainability constraints of the government to (8), which also has the virtue of not depending on the history of submechanisms up to that point.[26] Moreover, the lemma also shows that in any sustainable mechanism (8) is necessary.

Remark 1 By allowing all citizens to claim to be the worst type in the punishment phase, Assumption 3 plays a crucial role in the proof of Lemma 1. An alternative game form delivering Lemma 1 without Assumption 3 is as follows

1. At the beginning of period t, the government offers a menu of labor-consumption bundles, possibly dependent on past histories, denoted by $\mathcal{M}_t^C$.

[25]In particular, the following social plan (M, x) is a sustainable: $x_t = 0$ for all t and M_t always assigns zero labor (and consumption) to all reports.

[26]This statement refers to the sustainability constraint, (8). The optimal mechanism will clearly make allocations depend on the history of individual messages.

2. Individuals choose how much labor to supply, $l_t \in [0, \bar{l}]$, and production takes place.

3. The government decides $\xi_t \in \{0,1\}$. If $\xi_t = 1$, all output is expropriated, and consumed or distributed as before. If $\xi_t = 0$, production is distributed among agents who have chosen labor supply level as specified in the menu $\mathcal{M}_t^C$, and those who have chosen a labor supply level that is not in the menu (i.e., l_t such that $\nexists c_t \in \mathbb{R}_+$ with $(l_t, c_t) \in \mathcal{C}_t$), receive zero consumption.

This game form directly imposes "freedom of labor supply", meaning that it is individuals who decide how much labor to supply, whereas with our original game form, the submechanism determines how much labor individuals supply. With this game form, Lemma 1 applies exactly without Assumption 3. We chose the game form in the main text and Assumption 3 for two reasons: first, the game form in the main text is closer to the standard Mirrlees setup and the mechanism design structure; second, the alternative game form described here imposes an additional restriction on the set of mechanisms, such that individual allocations can differ only to the extent that individuals have chosen different labor-consumption bundles in the past. In contrast, our more general game form allows history-dependent allocations in which two individuals who have made different reports but have received the same consumption-labor bundle in the past may be treated differently in the future. Since they both lead to identical results, whether our baseline game form with Assumption 3 or this alternative game form without this assumption is preferred is a matter of taste, with no consequence for the rest of the results in the paper.

Remark 2 Yet another possible game form, which would lead to similar results, gives citizens the option to replace the government after a deviation. If replacement is costless, we would obtain a similar result to Lemma 1. See, for example, Acemoglu (2005b), Acemoglu, Golosov and Tsyvinski (2006).

Next, we define a *direct (sub)mechanism* as $M_t^* : \Theta^t \times H^t \to [0, \bar{l}] \times \mathbb{R}_+$. In other words, direct mechanisms involve a restricted message space, $Z_t = \Theta_t$, where individuals only report their current type.[27] We denote a strategy by the government inducing direct submechanisms along the equilibrium path by Γ^*.

Definition 3 *A strategy profile for the citizens, $\underline{\alpha}^*$, is truthful if, along the equilibrium path, we have that $\alpha_t^i \left(\theta^t \mid \theta^{t-1}, \rho^{t-1}, h^{t-1}\right) = \alpha^*$. We write $\underline{\alpha}^* = (\alpha^* \mid \alpha')$ to denote a truthful strategy profile.*

[27] Once again, more generally, this can be written as $M_t^* : \Theta^t \times H^t \to \Delta\left([0, \bar{l}] \times \mathbb{R}_+\right)$ to allow for stochastic mechanisms.

The notation $\underline{\alpha}^* = (\alpha^* \mid \alpha')$ emphasizes that individuals play truth-telling along the equilibrium path, but may play some different strategy profile, α', off the equilibrium path. Clearly, a truthful strategy against a direct mechanism simply amounts to reporting the true type of the agent. Let us next define $\underline{c}[\Gamma, \underline{\alpha}, \underline{h}]$, $l[\Gamma, \underline{\alpha}, \underline{h}]$ and $x[\Gamma, \underline{\alpha}, \underline{h}]$ as, respectively, the equilibrium consumption and labor supply distributions across individuals (as a function of the history of their reports), the sequence of government consumption levels resulting from the strategy profiles of the government and individuals, and the sequence of histories $\underline{h} = \{h^t\}_{t=0}^{\infty}$, such that all of these functions only condition or information available up to time t for allocations of time t.

Theorem 1 ***(Truthful Revelation Along the Equilibrium Path)*** *Suppose Assumptions 1-5 hold and that* Γ *and* $\underline{\alpha}$ *are a combination of strategy profiles that support a sustainable mechanism. Then, there exists another pair of equilibrium strategy profiles* Γ^* *and* $\underline{\alpha}^* = (\alpha^* \mid \alpha')$ *for some* α' *such that* Γ^* *induces direct submechanisms and* $\underline{\alpha}^*$ *induces truth telling along the equilibrium path, and moreover* $\underline{c}[\Gamma, \underline{\alpha}, \underline{h}] = \underline{c}[\Gamma^*, \underline{\alpha}^*, \underline{h}]$, $\underline{l}[\Gamma, \underline{\alpha}, \underline{h}] = \underline{l}[\Gamma^*, \underline{\alpha}^*, \underline{h}]$ *and* $x[\Gamma, \underline{\alpha}, \underline{h}] = x[\Gamma^*, \underline{\alpha}^*, \underline{h}]$.

Proof. See Appendix B. ■

The most important implication of this theorem is that for the rest of the analysis, we can restrict attention to truth-telling (direct) mechanisms on the side of the agents. The reason why, despite the lack of commitment and the self-interested preferences of the mechanism designer, a revelation principle type result holds is twofold: first, the government has a deviation within the same period; and second, individuals can use punishment strategies involving zero labor supply following a deviation by the government. The punishments strategies of citizens support a sustainable mechanism, making it the best response for the government to pursue the implicitly-agreed social plan (M, x). Given this sustainability, there is effective commitment on the side of the government *along the equilibrium path.* This notion is important to distinguish from the commitment that exists in the standard mechanism design problems where there is unconditional commitment (i.e., along all paths). In contrast, in our environment, there is no commitment *off the equilibrium path,* where the government can exploit the information it has gathered or expropriate part of the output. In fact, off the equilibrium path, non-truthful reporting by the individuals is important to ensure sustainability. Nevertheless, by definition, along the equilibrium path induced by a sustainable mechanism, the government prefers not to deviate from the implicitly-agreed social plan and thus individuals can report their types without the fear that this information or their labor supply will be misused.

3.3 The Best Sustainable Mechanism

Theorem 1 enables us to focus on direct mechanisms and truth-telling strategy α^* by all individuals. This implies that the best sustainable mechanism (and thus the best allocation) can be achieved by individuals simply reporting their types. Recall that at every date, there is an invariant distribution of θ denoted by $G(\theta)$. This implies that θ^t has an invariant distribution, which is simply the t-fold version of $G(\theta)$, $G^t(\theta)$ (since there is a continuum of individuals, each history θ^t occurs infinitely often).[28] Given this construction, we can write total labor supply given history h^t as $L_t(h^t) = \int_{\Theta^t} l_t(\theta^t, h^t)\, dG^t(\theta^t)$, and total consumption as $C_t(h^t) = \int_{\Theta^t} c_t(\theta^t, h^t)\, dG^t(\theta^t)$.[29] Moreover, since Theorem 1 establishes that any sustainable mechanism is equivalent to a direct mechanism with truth-telling on the side of the agents, we obtain the main result from the section, which will be used in the rest of the paper:

Proposition 1 *Suppose Assumptions 1-5 hold. Then, the best sustainable mechanism is a solution to the following maximization program:*

$$\mathbf{U}^{SM} = \max_{\{c_t(\theta^t,h^t), l_t(\theta^t,h^t), x_t(h^t), K_{t+1}(h^t)\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\left(\left(\theta^{i,t}, h^t\right)\right), l_t\left(\theta^{i,t}, h^t\right) \mid \theta_t^i\right)\right] \tag{9}$$

subject to some initial condition K_0, the resource constraint

$$K_{t+1}(h^t) = F\left(K_t(h^{t-1}), \int l_t(\theta^t, h^t)\, dG^t(\theta^t)\right) - \int c_t(\theta^t, h^t)\, dG^t(\theta^t) - x_t(h^t), \tag{10}$$

a set of incentive compatibility constraints for individuals,

$$\begin{aligned} &\mathbb{E}\left[\sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}\left(\theta^{i,t+s}, h^{t+s}\right), l_{t+s}\left(\theta^{i,t+s}, h^{t+s}\right) \mid \theta_{t+s}^i\right) \middle| \theta^{i,t}, h^t\right] \\ \geq\ &\mathbb{E}\left[\sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}\left(\hat{\theta}^{i,t+s}, h^{t+s}\right), l_{t+s}\left(\hat{\theta}^{i,t+s}, h^{t+s}\right) \mid \theta_{t+s}^i\right) \middle| \theta^{i,t}, h^t\right] \end{aligned} \tag{11}$$

for all t, all $\theta^{i,t} \in \Theta^t$, all $h^t \in H^t$ and all possible sequences of $\left\{\hat{\theta}_{t+s}^i\right\}_{s=0}^{\infty}$, and the sustainability constraint of the government

$$\mathbb{E}\left[\sum_{s=0}^{\infty} \delta^s v\left(x_{t+s}\left(h^{t+s}\right)\right) \middle| h^t\right] \geq v\left(F\left(K_t(h^{t-1}), \int l_t(\theta^t, h^t)\, dG^t(\theta^t)\right)\right), \tag{12}$$

for all t and $h^t \in H^t$.

[28] More formally, given the continuum of agents, we can apply a law of large numbers type argument, and each history θ^t will have positive measure. See, for example, Uhlig (1996).

[29] From now on, we suppress the $\tilde{}$'s to simplify notation and simply use c_t, l_t and x_t. Note also that $\int_{\Theta^t}$ here denotes Lebesgue integrals, and in what follows, we will suppress the range of integration, Θ^t.

Proof. The proof follows from Lemma 1 and Theorem 1. Suppose there exists an equilibrium $(\underline{\alpha}^{**}, \Gamma^{**})$, that maximizes (9). By the argument in the text, $(\underline{\alpha}^{**}, \Gamma^{**})$ will not feature $\xi_t = 1$ for any t. Therefore, $(\underline{\alpha}^{**}, \Gamma^{**})$ features a sequence of submechanisms $\left\{\hat{M}_t\right\}_{t=0}^{\infty}$, (potentially stochastic) consumption levels for the government, $\left\{\hat{x}_t\left(h^t\right)\right\}_{t=0}^{\infty}$ and $\xi_t\left(h^t\right) = 0$ for all t and $h^t \in H^t$. Then setting $(M, x) = \left(\left\{\hat{M}_t\right\}_{t=0}^{\infty}, \left\{\hat{x}_t\left(h^t\right)\right\}_{t=0}^{\infty}\right)$ implies that $(\underline{\alpha}^{**}, \Gamma^{**})$ support a sustainable mechanism. Then, use Theorem 1 to find $(\underline{\alpha}^{*}, \Gamma^{*})$ corresponding to a sustainable direct mechanism. This direct mechanism has to satisfy the resource constraint, (10), the incentive compatibility constraints of individuals at all dates, which instead of (6) can be written as (11) since Γ^* induces direct mechanisms. Finally, from Lemma 1, the constraint (12) ensures that Γ^* is a best response to citizens' strategies, $\underline{\alpha}^*$. ■

It is important to note that the objective function in the optimization only incorporates expectation at time $t = 0$ before individuals know their type (i.e., "behind the veil of ignorance"), while the incentive compatibility constraints, (11), require that there are no profitable deviations given any sequence of individual types and public histories. Note further that the maximization in (9) is over sequences $\left\{c_t\left(\theta^t, h^t\right), l_t\left(\theta^t, h^t\right), x_t\left(h^t\right), K_{t+1}\left(h^t\right)\right\}_{t=0}^{\infty}$, which allows potential randomization conditional on the realizations of the publicly observable aggregate variable h_t. Finally, this problem also defines $\mathbf{U}^{SM}$ as the ex ante value of the best sustainable mechanism for an individual.

The role of Theorem 1 in this formulation is obvious, since it enables us to write the program for the best sustainable mechanism as a direct mechanism with truth-telling, thus reducing the larger set of incentive compatibility constraints of individuals to (11).[30]

4 Sustainable Mechanisms and the Quasi-Mirrlees Program

Let us next define the *dynamic Mirrlees program* (with full-commitment, benevolent government and exogenous government expenditures). Imagine the economy needs to finance a (potentially stochastic) exogenous government expenditure $X_t\left(h^t\right) \geq 0$ at time t. Then the dynamic Mirrlees program of maximizing the time $t = 0$ (*ex ante*) utility of a representative agent, can be written as (e.g., Golosov, Kocherlakota and Tsyvinski, 2003, Kocherlakota, 2005):

$$\max_{\left\{c_t\left(\theta^t, h^t\right), l_t\left(\theta^t, h^t\right)\right\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\left(\theta^{i,t}, h^t\right), l_t\left(\theta^{i,t}, h^t\right) \mid \theta_t^i\right)\right]$$

[30] The equations in (11) focus on the incentive compatibility constraints that apply along the equilibrium path (expectations on both sides of the constraints are taken conditional on $\theta^{i,t}$). This is without any loss of generality, since (11) needs to hold for any sequence of reports $\left\{\hat{\theta}_{t+s}^i\right\}_{s=0}^{\infty}$, thus any potential deviation from time $t = 0$ is covered by this set of constraints.

subject to the incentive compatibility constraints, (11), and $C_t(h^t) + X_t(h^t) + K_{t+1}(h^t) \leq F(K_t(h^{t-1}), L_t(h^t))$ for all $h^t \in H^t$, where $C_t(h^t) = \int c_t(\theta^t, h^t)\, dG(\theta^t)$ and $L_t(h^t) = \int l_t(\theta^t, h^t)\, dG(\theta^t)$, and $c_t(\theta^t, h^t)$ and $l_t(\theta^t, h^t)$ are $\mathcal{F}_t$-measurable (see footnote 12). Moreover, we can add the feasibility constraint that $\{X_t(h^t)\}_{t=0}^{\infty}$ should be such that

$$\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \Lambda^{\infty} \text{ for all } h^t \in H^t,$$

where

$$\begin{aligned} \Lambda^{\infty} = \{ \quad & \{C_t, L_t\}_{t=0}^{\infty} \text{ such that } \exists \left\{c_t(\theta^t), l_t(\theta^t)\right\}_{t=0}^{\infty} \text{ satisfying (11)}, \\ & C_t(h^t) = \int c_t(\theta^t, h^t)\, dG(\theta^t), \text{ and } L_t(h^t) = \int l_t(\theta^t, h^t)\, dG(\theta^t)\}. \end{aligned} \tag{13}$$

In other words, for certain government expenditure sequences, $\{X_t\}_{t=0}^{\infty}$'s, the constraint set of this Mirrlees maximization problem can be empty (e.g., if $C_t = 0$ and $L_t > 0$, the incentive compatibility constraints of individuals cannot be satisfied). Thus it is important to ensure that $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \Lambda^{\infty}$ for all $h^t \in H^t$.

For a sequence $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \Lambda^{\infty}$ for all $h^t \in H^t$, we can define the *quasi-Mirrlees program* as

$$\mathcal{U}(\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty}) \equiv \max_{\left\{c_t(\theta^t, h^t), l_t(\theta^t, h^t)\right\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t(\theta^{i,t}, h^t), l_t(\theta^{i,t}, h^t) \mid \theta_t^i\right)\right] \tag{14}$$

subject to the incentive compatibility constraints, (11), and two additional constraints

$$\int c_t(\theta^t, h^t)\, dG(\theta^t) \leq C_t(h^t), \tag{15}$$

and

$$\int l_t(\theta^t, h^t)\, dG(\theta^t) \geq L_t(h^t), \tag{16}$$

for all $h^t \in H^t$. Clearly this program takes the sequence $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty}$ as given and maximizes the ex ante expected utility of an individual subject to the usual incentive compatibility constraints as well as two additional constraints. The first, (15), requires the sum of consumption levels across agents for all report histories to be no greater than some number C_t, while the second, (16), requires the sum of labor supplies to be no less than some amount L_t. The functional $\mathcal{U}(\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty})$ defines the maximum *ex ante* ($t = 0$) utility of an agent in this economy for a given sequence $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty}$.[31]

[31]We show in the Appendix that $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is a well-defined functional. Nevertheless, the incentive compatibility constraints embedded in (11) do not form a convex set. For this reason, in the Appendix, we follow Prescott and Townsend (1984a,b) and allow lotteries to convexify the constraint set and establish concavity of $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ in $\{C_t, L_t\}_{t=0}^{\infty}$. This will change the exact form of the optimization problem, but not its economic essence. For this reason, we relegate the formalism of the lotteries to Appendices C and D, and in the text, we assume that $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is concave. In the Appendix, we also prove that $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is differentiable, which we again assume to be the case in the text.

Returning to the dynamic Mirrlees program, for a given sequence of government expenditures $\{X_t(h^t)\}_{t=0}^{\infty}$, this can clearly be written as:

$$\max_{\{C_t(h^t), L_t(h^t), K_{t+1}(h^t)\}_{t=0}^{\infty}} \mathcal{U}(\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty}) \tag{17}$$

subject to a given level of K_0 and to

$$C_t(h^t) + X_t(h^t) + K_{t+1}(h^t) \leq F(K_t(h^{t-1}), L_t(h^t)), \text{ and } \{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \Lambda^{\infty}, \tag{18}$$

for all $h^t \in H^t$.

Therefore, we can represent the dynamic Mirrlees program as a solution to a two-step maximization problem, in which the first step is the quasi-Mirrlees formulation, yielding the functional $\mathcal{U}(\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty})$, and the second step is the maximization of $\mathcal{U}(\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty})$ over (potentially stochastic) sequences $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t)\}$ subject to a resource constrained and feasibility.

Now again using the quasi-Mirrlees formulation, the characterization of the best sustainable mechanism, (9), can be written as

$$\max_{\{C_t(h^t), L_t(h^t), x_t(h^t), K_t(h^t)\}_{t=0}^{\infty}} \mathcal{U}(\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty}) \tag{19}$$

subject to

$$C_t(h^t) + x_t(h^t) + K_{t+1}(h^t) \leq F(K_t(h^{t-1}), L_t(h^t)) \text{ and } \{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \Lambda^{\infty}, \tag{20}$$

for all $h^t \in H^t$, and also subject to (12). The only difference between the dynamic Mirrlees program in (17)-(18) and the best sustainable mechanism in (19)-(20)-(12) is the presence of the sustainability constraint for the government, (12), which also makes $\{x_t(h^t)\}_{t=0}^{\infty}$ and endogenously chosen sequence instead of the exogenously given $\{X_t\}_{t=0}^{\infty}$. This formulation establishes the following theorem.

Theorem 2 *Suppose Assumptions 1-5 hold. Then, the best sustainable mechanism solves a quasi-Mirrlees program for some sequence* $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \Lambda^{\infty}$.

Proof. This follows immediately from rewriting (9)-(12) from Proposition 1 as a two-step maximization program, and expressing (10) as $x_t(h^t) = F(K_t(h^{t-1}), L_t(h^t)) - C_t(h^t) - K_{t+1}(h^t)$. ■

Consequently, any allocation consistent with the best sustainable mechanism is a solution to a problem that maximizes the ex ante utility of the citizens. Despite the political economy constraints and the resources extracted by the government from the society, the mechanism

will maximize the ex ante utility of the citizens given some resource constraints (which are in addition to the resource constraints imposed by feasibility).[32] This theorem also enables us to represent the differences between the dynamic Mirrlees program and the best sustainable mechanism purely in terms of *aggregate distortions*, corresponding to what the sequences $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \Lambda^{\infty}$ are (and how they differ from the solution to the dynamic Mirrlees program in (17)-(18)).

To make more progress, we need to characterize the behavior of the sequences $\{C_t, L_t\}_{t=0}^{\infty}$ (and $\{x_t\}_{t=0}^{\infty}$) under the best sustainable mechanism, which is what we turn to next.

5 The Economy with Private Histories

The dynamic behavior of the optimal sustainable mechanism is determined by the need to provide dynamic incentives both to the government and to individual agents. As is well known (e.g. Green, 1987, Phelan and Townsend, 1991, or Atkeson and Lucas, 1992), the behavior of individual allocations can be very complicated even in the absence of sustainability constraints on the government. In order to highlight the effect of government sustainability constraints, we first consider mechanisms with *private histories*, i.e., where individual histories are not observed by the government.[33] The restriction to private histories is purely a heuristic device, useful in separating different parts of the analysis. Section 6 below will drop this assumption and characterize the best sustainable mechanism in the history-dependent case without any restrictions on potential mechanisms or strategies.

[32] It is also interesting to highlight which features of our environment are important for Theorem 2. For this purpose, consider an environment without capital and suppose that labor supply is equal to output. Suppose that the government can tax individuals differentially according to how much they produce, with the maximum amount that can be extracted from an individual supplying labor l as $\tilde{\eta}(l)$, where $\tilde{\eta}: [0,\bar{l}] \to [0,\bar{l}]$. In this case, using an analog of Lemma 1 and suppressing dependence on public histories, we have the sustainability constraint as:

$$\sum_{s=0}^{\infty} \delta^s v(x_{t+s}) \geq v\left(\int \tilde{\eta}(l(\theta^t))\, dG^t(\theta^t)\right),$$

where $l(\theta^t)$ is the labor supply of an individual with type history θ^t, and the term $\int \tilde{\eta}(l(\theta^t))\, dG^t(\theta^t)$ captures the maximum amount that the government can expropriate given the technological restriction embedded in the function $\tilde{\eta}(\cdot)$ and the distribution of types given by $G^t(\theta^t)$. Unless $\tilde{\eta}(\cdot)$ is a linear function, Theorem 2 does not apply, and there would be further distortions relative to our baseline analysis.

[33] Here the term "private histories" refers to the fact that past reports of individuals are not conditioned upon, and is not related to the public histories, h^t's, which designate the potential randomness of aggregate variables.

5.1 Best Sustainable Mechanism with Private Histories

To further simplify the notation in this case, let us also assume that there is no capital, so that the aggregate production function of the economy is simply

$$Y_t = L_t, \tag{21}$$

where $K_0 = 0$ and L_t denotes the aggregate labor supply at time t.

The restriction to private histories implies that in admissible mechanisms, allocations must depend only on agents' current report (and potentially on public histories, h^t). In such an environment the incentive compatibility constraints for agents can be separated across time periods, and written as

$$u\left(c_t\left(\theta_t, h^t\right), l_t\left(\theta_t, h^t\right) \mid \theta_t\right) \geq u\left(c_t\left(\hat{\theta}_t, h^t\right), l_t\left(\hat{\theta}_t, h^t\right) \mid \theta_t\right) \tag{22}$$

for all $\hat{\theta}_t \in \Theta$ and $\theta_t \in \Theta$, and for all t and $h^t \in H^t$. Moreover, given the single crossing property in Assumption 3, (22) can be reduced to a set of incentive compatibility constraints only for neighboring types. Since there are $N+1$ types in Θ, this implies that (22) is equivalent to N incentive compatibility constraints.[34] The best sustainable mechanism with private histories maximizes (9) subject to (12), (22) and the resource constraint

$$C_t\left(h^t\right) + x_t\left(h^t\right) \leq L_t\left(h^t\right), \tag{23}$$

Recall now the quasi-Mirrlees program defined above. It is straightforward to see that because of "private histories", the optimal allocations of (c_t, l_t) depend only on $C_t\left(h^t\right)$ and $L_t\left(h^t\right)$ and are independent of any $C_s\left(h^s\right)$, $L_s\left(h^s\right)$ with $s \neq t$. This implies that $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is time separable, i.e., $\mathcal{U}(\{C_t\left(h^t\right), L_t\left(h^t\right)\}_{t=0}^{\infty}) = \mathbb{E}\sum_{t=0}^{\infty} \beta^t U(C_t\left(h^t\right), L_t\left(h^t\right))$ for some real-valued differentiable function $U : \mathbb{R}_+^2 \to \mathbb{R}$. The program for the best sustainable mechanism, (19)-(20), therefore becomes:

$$\max_{\{C_t(h^t), L_t(h^t), x_t(h^t)\}_{t=0}^{\infty}} \mathbb{E}\sum_{t=0}^{\infty} \beta^t U(C_t\left(h^t\right), L_t\left(h^t\right)) \tag{24}$$

subject to the resource constraint, (23), and the sustainability constraint,

$$\mathbb{E}\left[\sum_{s=0}^{\infty} \delta^s v(x_{t+s}\left(h^{t+s}\right)) \mid h^t\right] \geq v(L_t\left(h^t\right)), \tag{25}$$

[34] More specifically, in pure strategy direct mechanisms, there will be $N(N+1)$ incentive compatibility constraints, and Assumption 2 makes sure that only N of those, i.e., those between neighboring types, where the higher type may want to misreport to be the next lower type, may be binding.

for all t and $h^t \in H^t$.

As before, this problem is well defined only for some (C, L). We denote the set of such (C, L) pairs, which is a simplified version of (13), by Λ, i.e., $\Lambda \equiv \{(C, L) : \exists (c(\theta), l(\theta))$ s.t. (15), (16) and (22) are satisfied$\}$. Also define $\bar{w} \equiv \max_{(C,L)\in\Lambda} v(L - C) / (1 - \delta)$. We impose:

Assumption 6 ***(sustainability)*** *There exists* $(\bar{C}, \bar{L}) \in \arg\max_{(C,L)\in\Lambda} v(L - C) / (1 - \delta)$, *such that* $v(\bar{L} - \bar{C}) / (1 - \delta) > v(\bar{L})$.

This assumption ensures that the highest discounted utility that can be given to the government, $\bar{w}$, is sufficient to satisfy its sustainability constraint (25). Clearly this assumption is satisfied if the discount factor of the government, δ, is sufficiently large.

We now introduce our key concept of "*aggregate distortions*". Since our objective is to compare the additional distortions created by the self-interested and time-inconsistent behavior of the government, aggregate distortions are defined relative to the dynamic Mirrlees allocations.

Definition 4 *In the model with no capital and with private histories, we say that the (potentially stochastic) sequence* $\{C_t(h^t), L_t(h^t), x_t(h^t)\}_{t=0}^{\infty}$ *induced by the best sustainable mechanism* Γ^* *is* undistorted *at* t' *if* $\left\{\hat{C}_t(h^t), \hat{L}_t(h^t)\right\}_{t=0}^{\infty}$ *is a solution to (17) subject to (18) with* $\{X_t(h^t)\}_{t=0}^{\infty} = \{x_t(h^t)\}_{t=0}^{\infty}$ *for all* $h^t \in H^t$ *and* $C_{t'}\left(h^{t'}\right) = \hat{C}_{t'}\left(h^{t'}\right)$, $\hat{L}_{t'}\left(h^{t'}\right) = L_{t'}\left(h^{t'}\right)$ *for all* $h^{t'} \in H^{t'}$. *We say that* $\{C_t(h^t), L_t(h^t), x_t(h^t)\}_{t=0}^{\infty}$ *is* asymptotically undistorted, *if it is (almost surely) undistorted as* $t \to \infty$

This is a natural definition. It requires that the aggregate allocations in the best sustainable mechanism coincide with the allocations in the dynamic Mirrlees program where government expenditures are equal to what is being paid to the government under the best sustainable mechanism.[35] It can be noted that if the ruler could commit to a sequence of consumption levels $\{x_t(h^t)\}_{t=0}^{\infty}$ at time $t = 0$, then Definition 4 implies that the resulting allocation would always be undistorted. The question is whether when the sequence $\{x_t(h^t)\}_{t=0}^{\infty}$ is determined to satisfy the sustainability constraint of the government, (12), there will be aggregate distortions.

Definition 4 is general enough to cover the case in which $(C, L) \notin \mathrm{Int}\Lambda$ (where $\mathrm{Int}\Lambda$ denotes the interior of Λ). For the usual case where $(C, L) \in \mathrm{Int}\Lambda$, we have a simple condition characterizing undistorted allocations. In particular, since $U(C, L)$ is differentiable (see Appendix B), when $(C, L) \in \mathrm{Int}\Lambda$ the solution to the dynamic (full-commitment) Mirrlees program (17)-(18) satisfies:

$$U_C(C, L) = -U_L(C, L), \tag{26}$$

[35] Note also that if an allocation is distorted, then in the absence of the sustainability constraints it could be modified to create a Pareto improvement.

where U_C and U_L are the partial derivatives of $U(C, L)$ with respect to C and L. Condition (26) is intuitive: it requires the marginal cost of increasing output by one unit to be equal to the marginal benefit of doing so. The next proposition illustrates the relationship between distortions and this condition more explicitly (suppressing dependence on h^t to simplify notation):

Proposition 2 *Suppose Assumptions 1-2 hold. Consider a sequence of* $\{C_t, L_t\}_{t=0}^{\infty}$*. Then:*

1. *the marginal labor tax rate on the highest type of agent,* θ_N*, at time t is given by* $\tau_{N,t} = 1 + U_L(C_t, L_t) / U_C(C_t, L_t)$.
2. *if* $\{C_t, L_t\}_{t=0}^{\infty}$ *is undistorted at t, the labor supply decision of the highest type of agent is undistorted, i.e.,* $u_c(c_t(\theta_N), l_t(\theta_N) \mid \theta_N) = -u_l(c_t(\theta_N), l_t(\theta_N) \mid \theta_N)$.

Proof. Assumption 2 implies that we only need to check incentive compatibility constraints for neighboring types. Appendix C establishes that Lagrange multipliers exist. Let u_c and u_l be the partial derivatives of u (which exist by Assumption 1). Therefore, we have

$$\begin{aligned} u_c(c_t(\theta_N), l_t(\theta_N) \mid \theta_N)(1 + \lambda_{Nt}) &= \nu_{Ct}, \\ u_l(c_t(\theta_N), l_t(\theta_N) \mid \theta_N)(1 + \lambda_{Nt}) &= -\nu_{Lt}, \end{aligned}$$

where λ_{Nt} is the multiplier on incentive compatibility constraint between types θ_N and θ_{N-1} at time t, ν_{Ct} is the multiplier on (15) at t and ν_{Lt} is the multiplier on (16) at t. By the differentiability of $U(C, L)$ and the definition of Lagrange multipliers, $\nu_{Ct} = U_C(C_t, L_t)$ and $\nu_{Lt} = -U_L(C_t, L_t)$. Combining these equations, we have

$$-\frac{u_l(c_t(\theta_N), l_t(\theta_N) \mid \theta_N)}{u_c(c_t(\theta_N), l_t(\theta_N) \mid \theta_N)} = (1 - \tau_{N,t}) = -\frac{U_L(C_t, L_t)}{U_C(C_t, L_t)},$$

where the first equality defines $\tau_{N,t}$, and the second equality establishes the first part of the lemma. The second result follows immediately from setting $U_L(C_t, L_t) = -U_C(C_t, L_t)$ from the definition of an undistorted sequence, in particular, equation (26). ■

Let us next follow Thomas and Worrall (1990) and consider the recursive characterization of the best sustainable mechanism. By standard arguments (and again suppressing dependence on h^t), any solution to the following recursive maximization problem is a sustainable mechanism (see Appendix C):

$$V(w) = \max_{C, L, x, w'} \{U(C, L) + \beta V(w')\} \tag{27}$$

subject to

$$C + x \leq L,$$

$$w = v(x) + \delta w', \tag{28}$$

$$v(x) + \delta w' \geq v(L), \tag{29}$$

$$w' \in \mathbb{W} \text{ and } (C, L) \in \Lambda. \tag{30}$$

where w is the future utility promised to the government, $\mathbb{W}$ is the set of feasible values for w, and the requirement that $(C, L) \in \Lambda$ makes sure that we only look at feasible levels of aggregate consumption and labor supply. The program in (27) determines optimal policies for a given level of promised utility w. Once the value function $V(w)$ is determined from this program, the problem of finding the best sustainable mechanism is simply equivalent to choosing the initial value of promised utility to government, w_0, such that $w_0 \in \arg\max_w V(w)$.[36]

This program also makes the role of the sustainability constraint (29) clear. If the society wishes to produce more output (or supply more labor L), it can only do so by providing greater consumption to the government either today or in the future. Therefore, when this constraint is binding, the *social cost* of increasing output will be *greater* than U_L, thus leading to aggregate distortions relative to the dynamic Mirrlees benchmark.

The main result of this section is the following theorem:

Theorem 3 *Consider the economy with no capital and with private histories and suppose that Assumptions 1-3, 5 and 6 hold.*

1. *At $t = 0$, there is an aggregate distortion.*

2. *Suppose that $\beta \leq \delta$. Let Γ^* be the best sustainable mechanism inducing a sequence of values $\{w_t(h^t)\}_{t=0}^{\infty}$. Then $\{w_t(h^t)\}_{t=0}^{\infty}$ is a non-decreasing stochastic sequence in the sense that $w_{t+1}(h^{t+1}) \geq w_t(h^t)$ for all $h^{t+1} \in H^{t+1}$. Moreover, a steady state exists in that $\{w_t(h^t)\}_{t=0}^{\infty}$ converges (almost surely) to some $w^* \in [0, \bar{w}]$ and $\{C_t(h^t), L_t(h^t), x_t(h^t)\}_{t=0}^{\infty}$ converges (almost surely) to some (C^*, L^*, x^*), which is asymptotically undistorted.*

3. *If $\beta > \delta$, then aggregate distortions do not disappear even asymptotically.*

Proof. See Appendix C. ■

[36] There is a number of technical details related to this program. First, we have that $\mathbb{W} = [0, \bar{w}]$ where $\bar{w}$ is the maximal feasible and sustainable promised utility to the government, defined above. Moreover, in Appendix C we show that there may be room for improving on this program by randomizing over the values of w', and thus over C and L (i.e., considering lotteries for the government in the same way as we do for individuals). We relegate the discussion of this issue to Appendix C, but here we take the sequence of values given (promised) to the government $\{w_t\}_{t=0}^{\infty}$ as a stochastic process, with each element taking values from the set $\mathbb{W}$, and consequently, the sequence $\{C_t, L_t\}_{t=0}^{\infty}$ that results from the best sustainable mechanism is also a stochastic sequence. Finally, in the text, we assume that V is concave and differentiable, which are both proved in Appendix C.

Part 1 reflects the additional aggregate distortions arising from the self-interested behavior of the government. In particular, at $t = 0$, there is necessarily a positive distortion, reducing aggregate output. Intuitively, this distortion results from the fact that as output increases, the sustainability constraint (29) implies that more has to be given to the government, and this makes the effective cost of production higher for the agents. Consequently, the best sustainable mechanism induces an aggregate distortion, reducing the levels of aggregate labor supply and production below those that would arise in the dynamic Mirrlees program.

The most important results are in parts 2 and 3. Part 2 states that as long as $\beta \leq \delta$, asymptotically the economy converges to an equilibrium where there are no aggregate distortions and the marginal tax rate on the highest type is equal to zero. Therefore, this theorem, in combination with Theorem 2, implies that despite the political economy constraints and the commitment problems, many of the insights of the optimal taxation literature inspired by Mirrlees (1971) will continue to hold. Consequently, when the government is at least as patient as the citizens, lessons from the optimal taxation literature are not only normative, but may also help us understand how tax systems are designed in practice where politicians are motivated by their own objectives, such as self-enrichment or reelection.

Part 3 of the theorem is perhaps more important. This part states that if the government is less patient than the agents, distortions will not disappear. Since in many realistic political economy models, the government or politicians are more short-sighted than citizens, this part of the theorem may imply that in a number of important cases, political economy considerations will lead to additional distortions that will not disappear even asymptotically.

We now give a heuristic argument to support this theorem (while the full proof is in Appendix B). Let γ and $\psi \geq 0$ be the Lagrange multipliers on the constraints (28) and (29) respectively. Lemma 10 in Appendix B shows that $V(w)$ is differentiable. Furthermore, in the text, we simplify the discussion by assuming that $(C, L) \in \text{Int}\Lambda$ and $w' \in \text{Int}\mathbb{W}$. Therefore, taking the first order condition with respect to w' and using the Envelope theorem, we obtain that

$$\frac{\beta}{\delta} V'(w') = -\psi - \gamma = V'(w) - \psi \tag{31}$$

The other first order conditions with respect to C, L and x imply:

$$U_C + U_L = \psi v'(L), \text{ and} \tag{32}$$

$$v'(x)(\psi + \gamma) = U_C. \tag{33}$$

Equation (32) makes it clear that aggregate distortions are related to ψ. It is also evident that we must have $\psi > 0$ at $t = 0$, otherwise the government would receive $w_0 = 0$ initially;

which together with the sustainability constraint (29) would imply $C_t = L_t = 0$ for all t, and this cannot be a solution when $\psi = 0$. Equation (32) then yields that $U_C + U_L > 0$, and Proposition 2 implies that the marginal tax rate on the highest type will be strictly positive, i.e., $\tau_N > 0$ at $t = 0$.

Part 2 of Theorem 3 states that, as long as $\beta \leq \delta$, eventually aggregate distortions will disappear and marginal labor income taxes on the highest type will tend to zero. In many ways, this is a surprising result, but the structure of the model makes the intuition clear. To see why, let us start with the case where $\beta = \delta$, in which case equation (31) implies

$$V'(w') = V'(w) - \psi \leq V'(w).$$

The inequality above is strict when the sustainability constraint on the government (29) binds. This, combined with the concavity of the value function $V(\cdot)$, which is proved in Lemma 8 in Appendix B, implies that $w' \geq w$, with $w' > w$ if $\psi > 0$ and $w' = w$ if $\psi = 0$. This shows that the promised utilities for the government are nondecreasing as stated in part 2 of Theorem 3.

The intuition for why the rewards to the government are increasing (nondecreasing) is as follows. The incentives for the government in the current period are provided by both consumption in the current period, x, and by consumption in future periods represented by the promised utility w. Therefore, future government consumption not only relaxes the sustainability constraint in the future, but also in all prior periods. Thus, all else equal, optimal incentives for government are backloaded. The intuition for this backloaded compensation scheme is similar to the reasons why in principal-agent models backloading compensation may be useful (see, for example, Ray, 2002).

Since promised values to the government form a nondecreasing sequence $\left\{w_t\left(h^t\right)\right\}_{t=0}^{\infty}$ and are in a compact set $\mathbb{W} = [0, \bar{w}]$, they will converge to some value w^*. If $w^* < \bar{w}$, (31) immideately implies that $\psi = 0$. In other words, the Lagrange multiplier on the sustainability constraint of the government, (29), eventually reaches zero, and at this point, aggregate distortions disappear. Proposition 2 then implies that, as long as $(C, L) \in \text{Int}\Lambda$, the marginal tax rate on the labor supply of the highest type, θ_N, also vanishes as in the standard Mirrlees program.

The intuition for why the multiplier on the sustainability constraint eventually reaches zero is related to the fact that promised utilities to the government are increasing. Loosely speaking, we can remove the sustainability constraints in the very far future, without influencing the sequence of utilities promised to the government at time $t = 0$. This implies that eventually the multipliers on these sustainability constraints must tend to zero.

Finally, let us consider the case with $\delta < \beta$. Since government is less patient than the agents, backloading incentives for government becomes costly for agents. Consider any w for which constraint (29) does not bind. Then (31) implies that $V'(w') > V'(w)$, and thus $w' < w$, so that promised utilities will be decreasing when the sustainability constraint, (29), is slack. In fact, if a steady state (C^*, L^*, x^*) is ever reached, it will solve the following system of equations

$$1 + \frac{U_L}{U_C} = \left(1 - \frac{\delta}{\beta}\right) \frac{v'(L^*)}{v'(x^*)} \tag{34}$$

$$C^* + x^* = L^* \text{ and } v(x^*) = (1-\delta)v(L^*),$$

with the steady-state utility of the government equal to $w^* = v(L^*)$. Equation (34) immediately shows that when $\delta < \beta$, were a steady state reached, it would feature a positive labor distortion, as claimed in part 3 of Theorem 3. The intuition for the presence of (asymptotic) aggregate distortions in this case is directly related to the fact that when the government is less patient than the agents, backloading does not work. Since backloading was essential for the multiplier on the sustainability constraint (29) going to zero, this multiplier remains positive, and the additional distortions created by the sustainability constraint remain even asymptotically.

5.2 An Example for an Economy with Private Histories and No Capital

We now illustrate the results from the previous subsection with a simple numerical example. Consider an economy with two types, i.e., $\Theta = \{\theta_0, \theta_1\}$ and

$$u(c, l \mid \theta) = \sqrt{c} - \frac{l^2}{m\theta}, \tag{35}$$

where m is a parameter determining the relative disutility of labor. We continue to assume that type θ_0 is disabled and cannot supply any labor, so $\theta_0 = 0$, and we normalize $\theta_1 = 1$. Let us also assume that a fraction $\pi = 1/2$ of the population is of type θ_1 and that the utility function of the government is also given by $v(x) = \sqrt{x}$.

Since type θ_0 cannot supply any labor, we have $l(\theta_1) = L/\pi$. Moreover, the incentive compatibility constraint for type θ_1 is

$$\sqrt{c(\theta_1)} - \frac{l(\theta_1)^2}{m\theta_1} \geq \sqrt{c(\theta_0)}. \tag{36}$$

Then $c(\theta_0)$ and $c(\theta_1)$ can be determined as solutions to (36) holding as equality and to the resource constraint, $(1-\pi)c(\theta_0) + \pi c(\theta_1) = C$. Given this structure, $U(C, L)$ can be computed directly and used with the program (27) to derive the value function $V(w)$.

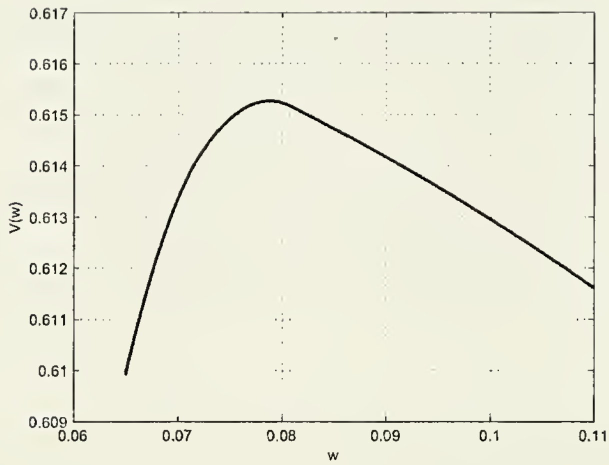

Figure 1: Theorem 3 part 2. Value function with $\beta = \delta = 0.9$.

We take a baseline case where the government is as patient as the citizens, $\beta = \delta = 0.9$ and $m = 5$. In Figure 1 we show the shape of the value function.[37] Note that $V(w)$ is inverse U-shaped. The increasing part is due to the fact that if the government is given too low a level of utility, the sustainability constraint will force the economy to produce a very low level of output. For this reason, the relevant part of the value function $V(w)$ is the segment after the peak, which is everywhere decreasing. In fact, as noted above, with the best sustainable mechanism, the economy will start at the peak of the function $V(w)$.

Next, we show the time path of normalized promised values to the government (defined as $(1-\delta)w$) and aggregate distortions both for the baseline case, $\delta = 0.9$, and also for a range of lower discount factors, $\delta = 0.8$, 0.7, and 0.6. Figure 2 plots the time path of the promised value to the government, w, for these four different cases. The lowest curves is for $\delta = 0.6$, and then, respectively, for $\delta = 0.7$, 0.8 and 0.9. Consistent with the results in Theorem 3 part 2, when $\delta = \beta$, $\{w_t\}$ is an increasing sequence and converges to some level w^*. Interestingly, in these examples the sequence $\{w_t\}$ is everywhere increasing even when $\delta < \beta$.

Figure 3, in turn, depicts the evolution of the aggregate distortion, $1 + U_L/U_C$ (which, from Proposition 2, is also equivalent to the marginal tax on type θ_1). The lowest curve shows the case where $\beta = \delta$, and consistent with part 2 of Theorem 3, the aggregate distortion converges to zero. An interesting feature of the example is that the convergence of $\{w_t\}$ and of distortions

[37]We compute everything non-recursively, assuming that a steady state is reached after T periods and then solving a sequence problem, which turns out to be faster and more accurate than the recursive approach in the computations.

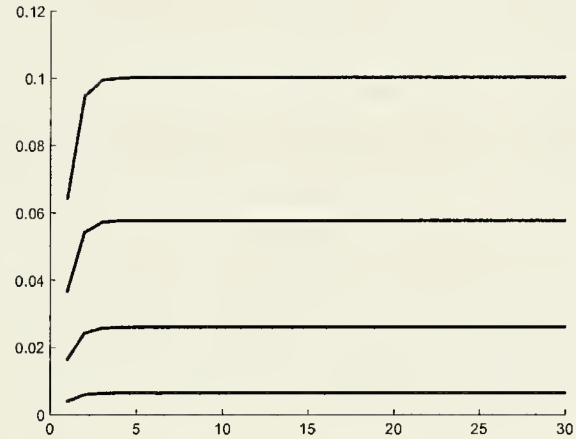

Figure 2: Time path of normalized promised values to the government, $\{(1-\delta)\,w_t\}$, with $\beta = 0.9$ and $\delta = 0.9; 0.8; 0.7; 0.6$. Higher curves correspond to higher values of δ.

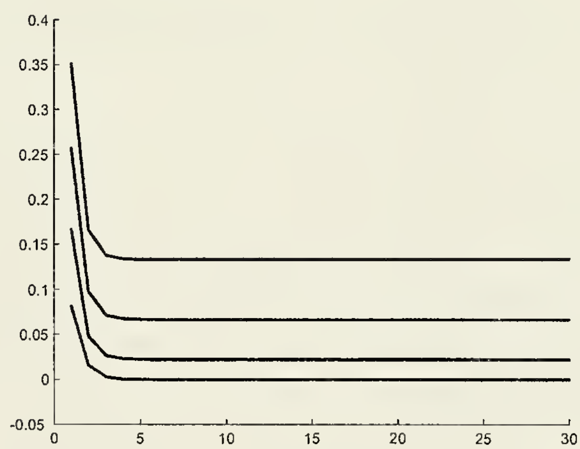

Figure 3: Time path of distortions with $\beta = 0.9$ and $\delta = 0.6, 0.7, 0.8, 0.9$. Higher curves correspond to lower values of δ.

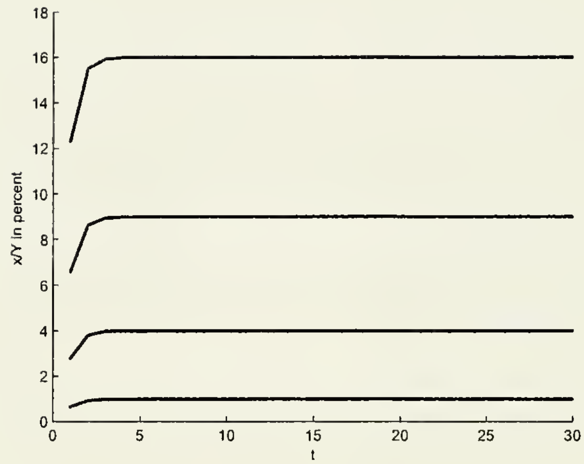

Figure 4: Time path of x_t/Y_t with $\beta = 0.9$ and $\delta = 0.6, 0.7, 0.8, 0.9$. Higher curves correspond to lower values of δ.

is rather fast. This suggests that the best sustainable mechanism may converge to a mechanism without aggregate distortions and with zero marginal tax rate on the highest type agents very rapidly.

The figure also shows that, as predicted by part 3 of Theorem 3 and in contrast to the case with $\beta = \delta$, when $\delta < \beta$ aggregate distortions do not disappear asymptotically; in fact, they could be quite sizable. For example, when $\delta = 0.6$, the aggregate distortion converges to an asymptotic value of 0.15 (the highest curve in the graph).

A final interesting feature is how much of the economy's output is being captured by the government. Figure 4 shows this again for $\delta = 0.9$, 0.8, 0.7, and 0.6. When the discount factor of the government is equal to that of the citizens, even in the asymptotic equilibrium, the government receives a very small fraction of the output (and $w^* < \bar{w}$ with the notation in the previous subsection). As we consider lower discount factors for the government, its temptation to deviate increases, so a higher fraction of the output goes to the government, but even with $\delta = 0.6$, this is only 16% of total output.

6 Optimal History-Dependent Sustainable Mechanisms

6.1 Characterization of Best Sustainable Mechanism

We now return to the analysis of the general problem in Sections 2-4 without the restriction to private histories, and we also re-incorporate capital. Despite the fact that individual allocations

will now be a function of the entire history of reports by the individual, the analysis will parallel the discussion of the best sustainable mechanism with private histories in Section 5 (and from Theorem 1, individuals will continue to report their types truthfully).

The quasi-Mirrlees program was defined above in (14), and Theorem 2 established that the best sustainable mechanism solves a quasi-Mirrlees program. In addition, Appendix C shows that $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is differentiable in the sequences $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$. This implies that we can think of variations in sequences $\{C_t, L_t\}_{t=0}^{\infty}$ where only one element, C_s or L_s for some specific s is varied. We denote the derivative of $\mathcal{U}$ with respect to such variations by $\mathcal{U}_{C_s}(\{C_t, L_t\}_{t=0}^{\infty})$ and $\mathcal{U}_{L_s}(\{C_t, L_t\}_{t=0}^{\infty})$ or simply by $\mathcal{U}_{C_s}$ and $\mathcal{U}_{L_s}$. We also denote the partial derivatives of the production function with respect to labor and capital at time s by F_{L_s} and F_{K_s}. We next extend our definition of an undistorted equilibrium to this more general environment:

Definition 5 *We say that the (potentially stochastic) sequence* $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t), x_t(h^t)\}_{t=0}^{\infty}$ *induced by the best sustainable mechanism* Γ^* *is* undistorted *at* t' *if* $\left\{\hat{C}_t(h^t), \hat{L}_t(h^t), \hat{K}_{t+1}(h^t)\right\}_{t=0}^{\infty}$ *is a solution to (17) subject to (18) with* $\{X_t(h^t)\}_{t=0}^{\infty} = \{x_t(h^t)\}_{t=0}^{\infty}$ *for all* $h^t \in H^t$ *and* $C_{t'}\left(h^{t'}\right) = \hat{C}_{t'}\left(h^{t'}\right)$, $\hat{L}_{t'}\left(h^{t'}\right) = L_{t'}\left(h^{t'}\right)$, $\hat{K}_{t'+1}\left(h^{t'}\right) = K_{t'+1}\left(h^{t'}\right)$ *for all* $h^{t'} \in H^{t'}$. *We say that* $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t), x_t(h^t)\}_{t=0}^{\infty}$ *is* asymptotically undistorted *if it is (almost surely) undistorted as* $t \to \infty$.

This definition is the natural generalization of Definition 4 to this more general environment. As in Definition 4, when $\{C_t, L_t\}_{t=0}^{\infty} \in \mathrm{Int}\Lambda^{\infty}$, the current definition implies that an undistorted allocation will satisfy

$$\mathcal{U}_{C_t} \cdot F_{L_t} = -\mathcal{U}_{L_t} \text{ and } F_{K_{t+1}} \cdot \mathcal{U}_{C_{t+1}} = \mathcal{U}_{C_t} \tag{37}$$

at time t (or as $t \to \infty$). Here, the first condition requires the marginal cost of effort at time t given the utility function $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ to be equal to the increase in output from the additional effort times the marginal utility of additional consumption, and is thus an immediate generalization of condition (26) from the economy without capital. The second one requires the cost of a decline in the utility by saving one more unit to be equal to the increase in output in the next period times the marginal utility of consumption then. Once again, these are aggregate conditions since they are defined in terms of the utility functional $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$, which represents the ex ante maximal utility of an individual subject to incentive constraints. Moreover, if a steady state exists and the conditions in (37) hold as $t \to \infty$, then it is also clear that $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t), x_t(h^t)\}_{t=0}^{\infty}$ must be undistorted. We therefore say that there are no asymptotic aggregate distortions on capital accumulation (or no aggregate capital

taxation) if $F_{K_{t+1}} \cdot \mathcal{U}_{C_{t+1}} = \mathcal{U}_{C_t}$ and no aggregate distortions on labor supply if $\mathcal{U}_{C_t} \cdot F_{L_t} = -\mathcal{U}_{L_t}$ as $t \to \infty$.[38]

Let us use the notation $\{C, L\} \in \bar{\Lambda}^\infty$ if $\{C_t(h^t), L_t(h^t)\}_{t=0}^\infty \in \Lambda^\infty$ and $C_t(h^t) \to C$ and $L_t(h^t) \to L$ almost surely, and the steady-state pair $\{C, L\}$ is asymptotically feasible.[39] When $\{C_t(h^t), L_t(h^t), K_t(h^t)\} \to (C^*, L^*, K^*)$ almost surely, define $\mathcal{U}^*_{C_t} = \mathcal{U}_{C_t}(\{C_t(h^t), L_t(h^t)\}_{t=0}^\infty)$. Finally, with analogy to the previous section, define the maximal steady-state level of utility for the ruler as: $\bar{w} \equiv \max_{\{C,L\} \in \bar{\Lambda}^\infty, K \geq 0} v(F(K, L) - C - K) / (1 - \delta)$. The key sustainability assumption is a generalization of Assumption 6 to this environment and requires that when the government receives a utility of $\bar{w}$, the sustainability constraint, (12), is satisfied:

Assumption 7 ***(general sustainability)*** *There exists* $(\bar{C}, \bar{L}, \bar{K}) \in \arg\max_{\{C,L\} \in \bar{\Lambda}^\infty, K \geq 0} v(F(K, L) - C - K) / (1 - \delta)$ *such that* $v(F(\bar{K}, \bar{L}) - \bar{C} - \bar{K}) / (1 - \delta) > v(F(\bar{K}, \bar{L}))$.

Then the main result of this section parallels Theorem 3, but is weaker in some respects:

Theorem 4 *Consider the model with the general environment and suppose that Assumptions 1-5 and 7 hold and that there exists* $\{C_t(h^t), L_t(h^t)\}_{t=0}^\infty \in Int\Lambda^\infty$ *(with* $L_t(h^t) > 0$ *for some* t *and all* $h^t \in \bar{H}^t$ *for some positive probability* $\bar{H}^t \subset H^t$*).*

1. *There exists* $t' < \infty$ *such that there are aggregate distortion on capital accumulation and labor supply at* t'.

Let Γ^* *be the best sustainable mechanism, inducing a sequence of consumption, labor supply and capital levels* $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t)\}$. *Suppose a steady state exists such that as* $t \to \infty$, $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t)\} \to (C^*, L^*, K^*)$ *almost surely. Moreover, let* $\varphi = \sup\{\varrho \in [0,1] : \text{plim}_{t\to\infty} \varrho^{-t} \mathcal{U}^*_{C_t} = 0\}$, *where* $\varphi < 1$. *Then:*

2. *If* $\varphi \leq \delta$, *then (almost surely) there are no asymptotic aggregate distortions on capital accumulation and labor supply.*

3. *If* $\varphi > \delta$, *then aggregate distortions on capital accumulation and labor supply do not disappear even asymptotically.*

[38]Notice that we no longer have the equivalent of Proposition 2, since without specifying the stochastic process for $\theta_{i,t}$, we cannot establish the equivalent of the single-crossing property in consumption sequences. This problem is common in models of dynamic taxation, even with the full commitment assumption, see, for example, Golosov, Kocherlakota and Tsyvinski (2003).

[39]Note that despite the similarity of the symbols, $\bar{\Lambda}^\infty$ and Λ^∞ are very different sets. Λ^∞ is a subset of the vector space of bounded infinite sequences, ℓ_∞, while $\bar{\Lambda}^\infty \subset \mathbb{R}^2$.

Proof. See Appendix D. ∎

The major results from Theorem 3 continue to hold here.[40] The most important difference is that instead of comparing the government discount factor, δ, to β, we now compare it to the rate at which the ex ante marginal utility of consumption $\mathcal{U}^*_{C_t}$ is declining in the steady state, denoted by φ. Clearly, in the case where $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is time separable as in Theorem 3, the rate at which $\mathcal{U}^*_{C_t}$ declines in steady state is exactly equal to β, so that the results in this theorem are closely related to those in Theorem 3. Moreover, in reality, φ is the "fundamental discount factor" of the citizens, since it measures how one unit of resources at time t compares with one unit of resources at time $t+1$ (from the viewpoint of $t = 0$). Only in special cases (e.g., without any dynamic incentive linkages) does this fundamental discount factor coincide with β. Therefore, the case of $\varphi \leq \delta$ indeed corresponds to the situation in which the government is as patient as or more patient than the citizens.

The first part of the theorem states that the sustainability constraint of the government, (12), necessarily introduces a distortion, though because of the ex ante utility function of the citizens, $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$, is nonseparable, we can no longer be sure that this distortion will be present at $t = 0$. The most important results are again contained in parts 2 and 3 of the theorem. Part 2 states that as long as a steady state exists and $\mathcal{U}^*_{C_t}$ declines sufficiently rapidly, the multiplier of the sustainability constraint goes to zero. This establishes that the sequence $\{C_t, L_t, K_t\}_{t=0}^{\infty}$ is asymptotically undistorted, with no aggregate labor supply and capital accumulation distortions. This generalizes the results from the economy with no capital and private histories to the much more general environment here. Part 3, on the other hand, states that if the discount factor of the government δ is sufficiently low, then aggregate distortions will not disappear, even asymptotically. The significance of this result is even greater than in Theorem 3, since it implies not only additional distortions on labor, but also *positive aggregate capital taxes* in contrast to the existing literature on dynamic fiscal policy.

Once again, we provide a heuristic argument here, leaving the proof to the Appendix. Since the objective function is no longer time separable, to characterize the best sustainable mechanism in this case, we follow Marcet and Marimon (1998) and form a Lagrangian of the form (again suppressing dependence on history h^t for notational simplicity):

$$\max_{\{C_t, L_t, K_t, x_t\}_{t=0}^{\infty}} \mathcal{L} = \mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) + \sum_{t=0}^{\infty} \delta^t \left\{\mu_t v(x_t) - (\mu_t - \mu_{t-1}) v(F(K_t, L_t))\right\} \tag{38}$$

[40] The parts that are missing from this theorem relative to Theorem 3 are that the aggregate distortion is at $t = 0$ (instead it could be at some later date), that the sequence of promised values to the government is increasing and a statement that a steady state exists. Moreover, as noted in footnote 38, there is no equivalent of Proposition 2 in this case. Finally, the theorem imposes the rather weak assumption that there exists $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \mathrm{Int}\Lambda^{\infty}$ with $L_t(h^t) > 0$ for some t and some h^t.

subject to

$$C_t + x_t + K_{t+1} \leq F(K_t, L_t), \text{ and} \tag{39}$$

$$\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty},$$

for all t, where $\mu_t = \mu_{t-1} + \psi_t$ with $\mu_{-1} = 0$ and $\delta^t \psi_t \geq 0$ is the Lagrange multiplier on the constraint (12).[41]

The differentiability of $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ implies that for $\{C_t, L_t\}_{t=0}^{\infty} \in \text{Int}\Lambda^{\infty}$, we have:[42]

$$\mathcal{U}_{L_t} - \delta^t(\mu_t - \mu_{t-1})v'(F(K_t, L_t))F_{L_t} = -\mathcal{U}_{C_t} \cdot F_{L_t} \tag{40}$$

$$\mathcal{U}_{C_t} = \left[\mathcal{U}_{C_{t+1}} - \delta^t(\mu_{t+1} - \mu_t)v'(F(K_{t+1}, L_{t+1}))\right] F_{K_{t+1}} \tag{41}$$

Since $\mu_t \geq \mu_{t-1}$, we obtain that

$$-\mathcal{U}_{L_t} \leq \mathcal{U}_{C_t} \cdot F_{L_t}, \text{ and} \tag{42}$$

$$\mathcal{U}_{C_t} \leq \mathcal{U}_{C_{t+1}} \cdot F_{K_{t+1}}. \tag{43}$$

These two conditions imply that there may be positive distortions in labor supply and capital accumulation—if the inequalities are strict, the marginal product of labor and capital would be too high (relative to the full commitment Mirrlees allocation).

The first-order condition with respect to x_t, on the other hand, yields:

$$\frac{\mathcal{U}_{C_t}}{\delta^t v'(x_t)} = \mu_t \leq \mu_{t+1} = \frac{\mathcal{U}_{C_{t+1}}}{\delta^{t+1} v'(x_{t+1})}. \tag{44}$$

By construction, μ_t is an increasing sequence, so it must either converge to some value μ^* or go to infinity. Suppose that (C_t, L_t, K_t) converges to some (C^*, L^*, K^*)—and x_t converges to

[41] To derive (38), form the Lagrangian

$$\mathcal{L}' = \mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) + \sum_{t=0}^{\infty} \delta^t \psi_t \left[\sum_{s=0}^{\infty} \delta^s v(x_{t+s}) - v(F(K_t, L_t))\right],$$

then note that

$$\sum_{t=0}^{\infty} \delta^t \psi_t \sum_{s=0}^{\infty} \delta^s v(x_{t+s}) = \sum_{t=0}^{\infty} \delta^t \mu_t v(x_t)$$

where $\mu_t = \mu_{t-1} + \psi_t$ with $\mu_{-1} = 0$. Substituting this in $\mathcal{L}'$ above gives (38).

[42] To obtain these equations, let the multiplier on constraint (39) at time t be κ_t. Then the first-order condition with respect to C_t gives $\mathcal{U}_{C_t} = \kappa_t$, while the first-order condition with respect to L_t gives

$$\mathcal{U}_{L_t} - \delta^t \left(\mu_t - \mu_{t-1}\right) v'(F(K_t, L_t))F_{L_t} = -\kappa_t F_{L_t}.$$

Substituting for κ_t gives (40). The first-order condition with respect to K_{t+1}, on the other hand, gives

$$-\delta^t(\mu_{t+1} - \mu_t)v'(F(K_{t+1}, L_{t+1}))F_{K_{t+1}} + \kappa_t - \kappa_{t+1}F_{K_{t+1}} = 0$$

Substituting for κ_t and κ_{t+1} and rearranging gives (41).

$x^* = L^* - C^* - K^*$. If U^*_{Ct} is proportional to some $\varphi \leq \delta$, then we can show that μ_t (almost surely) converges to some value $\mu^* < \infty$, and that both (42) and (43) must hold as equality (see the proof of Theorem 4), establishing the result stated in part 2 of the theorem. In contrast, if $\mathcal{U}^*_{Ct}$ is proportional to some $\varphi > \delta$, then μ_t tends to infinity and aggregate distortions do not disappear.

6.2 Example for History-Dependent Mechanisms

We now briefly illustrate the results of Theorem 4 and show how in some simple cases, φ defined as $\sup\{\varrho \in [0,1] : \text{plim}_{t\to\infty} \varrho^{-t} U^*_{Ct} = 0\}$ is again equivalent to the discount factor of the agents, β. In particular, let us consider the following economy without capital and with "almost constant types" as explained below. There are two types $\Theta = \{\theta_0, \theta_1\}$ and the utility function is

$$u(c, l \mid \theta) = u(c) - \chi(l/\theta),$$

where u is continuously differentiable, increasing and strictly concave and χ is continuously differentiable, increasing and strictly convex. Furthermore, suppose that u satisfies Inada-type conditions, so that first-order conditions are always satisfied as equality. We take $\theta_0 = 0$, so that the low type is again disabled and cannot supply any labor. Suppose that with probability π an individual is born as high type, and remains so with (iid) probability $1-\varepsilon$ in every period. With probability $1-\pi$, individual is born as low type, and remains low type forever. By almost constant types, we mean the limit of this economy as $\varepsilon \to 0$.[43] Then the quasi-Mirrlees formulation can be written as

$$\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) \equiv \max_{\{c_t(\theta_0), c_t(\theta_1), l_t(\theta_1)\}_{t=0}^{\infty}} \pi \sum_{t=0}^{\infty} \beta^t \left[u(c_t(\theta_1)) - \chi(l_t(\theta_1)/\theta_1)\right] + (1-\pi) \sum_{t=0}^{\infty} \beta^t \left[u(c_t(\theta_0))\right], \tag{45}$$

subject to $\pi c_t(\theta_1) + (1-\pi) c_t(\theta_0) \leq \pi l_t(\theta_1) - x_t$ for all t, and

$$\sum_{t=0}^{\infty} \beta^t \left[u(c_t(\theta_1)) - \chi(l_t(\theta_1)/\theta_1)\right] \geq \sum_{t=0}^{\infty} \beta^t \left[u(c_t(\theta_0))\right],$$

where $L_t = \pi l_t(\theta_1)$ and $C_t = L_t - x_t$. The first constraint is the resource constraint for each t, while the second constraint is the incentive compatibility constraint sufficient for the high type to reveal his identity given the presence of effective commitment along the equilibrium path. Because $\varepsilon \to 0$, we do not specify other incentive compatibility constraints. Assigning

[43] If instead we use the alternative game form outlined in Remark 1, this example would work with constant types rather than almost constant types.

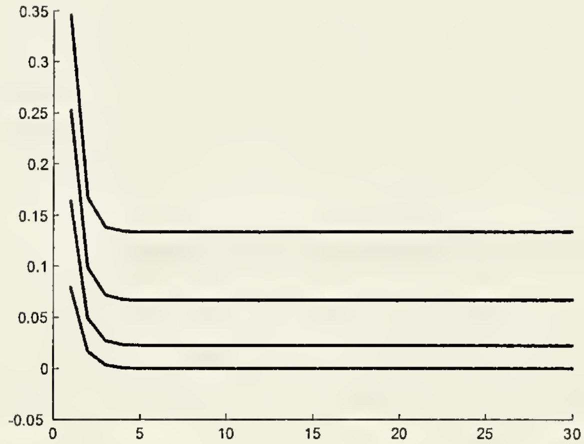

Figure 5: The time path of distortions for almost constant types with $\beta = 0.9$ and $\delta = 0.6, 0.7, 0.8, 0.9$. Higher curves correspond to lower values of δ.

Lagrange multipliers λ and $\beta^t \mu_t$ to these constraints, the first-order necessary conditions of this problem can be written as:

$$(\pi + \lambda) u' (c_t (\theta_1)) = \pi \mu_t \tag{46}$$

$$(1 - \pi - \lambda) u' (c_t (\theta_0)) = (1 - \pi) \mu_t, \tag{47}$$

and

$$\frac{(\pi + \lambda)}{\theta_1} \chi' (l_t (\theta_1) / \theta_1) = \mu_t. \tag{48}$$

Equations (46)-(47) imply that

$$\frac{u' (c_t (\theta_1))}{u' (c_t (\theta_0))} = \frac{(1 - \pi - \lambda)}{(\pi + \lambda)}.$$

Consequently, there is constant risk-sharing between the two types in all periods. Moreover, if a steady state exists, so that $x_t \to x^*$, (46)-(48) combined imply that $c_t (\theta_1) \to c^{1*}$, $c_t (\theta_0) \to c^{0*}$, and $l_t (\theta_1) \to l^*$, and hence $\mu_t \to \mu^*$. Consequently, in this case $\varphi = \beta$, so Theorem 4 applies in exactly the same form as Theorem 3. Therefore, in this particular case, the rate at which the derivative $\mathcal{U}^*_{C_t}$ declines is easy to determine, and it does so at the same rate as the discount factor of the citizens, i.e., $\varphi = \beta$. It is also straightforward to see that the same argument generalizes to the case where there are more than two types.[44]

[44]In fact, we conjecture that whenever there exists a stationary distribution of consumption among individuals, $\varphi = \beta$, though we have not been able to prove this conjecture.

For a numerical illustration, we again consider the utility function (35) with two types and the same parameter values from subsection 5.2. For brevity, we simply show the aggregate distortion, $1+\mathcal{U}_{L_t}/\mathcal{U}_{C_t}$, in Figure 5. Consistent with part 2 of Theorem 4, when $\beta=\delta=0.9$, the lowest curve shows that the aggregate distortion converges to zero and the convergence is again rather fast. Instead, when $\delta<\beta$, the aggregate distortion converges to a positive, and potentially large, asymptotic value.[45]

7 Benevolent Time-Inconsistent Governments

The analysis so far was simplified by the fact that the government was purely self-interested. Although this case is of relevance for many political economy applications, it is also important to understand how the results generalize to the case considered by Roberts (1984), Freixas, Guesnerie and Tirole (1985), or Bisin and Rampini (2005), where the government is still benevolent, but "time inconsistent", i.e., unable to commit to a full dynamic mechanism. To do this, we now consider a more general utility function for the government of the form:

$$\mathbb{E}_t \sum_{s=0}^{\infty} \delta^s \left[(1-a)\, v\left(x_{t+s}\right) + a \left(\mathbb{E}_{t+s} \int u\left(c_{t+s}, l_{t+s} \mid \theta^{t+s}\right) dG^{t+s}\left(\theta^{t+s}\right) \right) \right], \tag{49}$$

where the second term is the average (expected) utility of the citizens at time $t+s$ (with expectations based on information available at time $t+s$, which includes the public history h^{t+s}), and $a>0$ (the case with $a=0$ is identical to our analysis above). Therefore, this utility function is identical to that of a purely-self-interested government when $a=0$, and identical to the fully-benevolent case when $\delta=\beta$ and $a=1$.

In this case, we need to strengthen Assumption 1 and assume separable utility, which is a standard assumption in most analyses of dynamic taxation (e.g., Golosov, Kocherlakota and Tsyvinski, 2003, Kocherlakota, 2005).

Assumption 1' (separable utility) $u(c,l \mid \theta) = u(c) - \chi(l \mid \theta)$, *where* $u : \mathbb{R}_+ \to \mathbb{R}$ *is continuously differentiable, strictly increasing and concave, and* $\chi(\cdot \mid \theta)$ *is continuously differentiable, strictly increasing and convex for all* $\theta \in \Theta$, *and satisfies* $\chi(0 \mid \theta) = 0$ *for all* $\theta \in \Theta$.

The next theorem shows that Theorem 1 and Proposition 1 continue to hold in this more general environment, but with Assumption 1' replacing Assumption 1:

[45] Note that the distortions are the same here as for the case with the private histories studied in subsection 5.2. The reason is that without the sustainability constraints, the problems would have the same solution. When sustainability constraints are present but satisfied, there are no other reasons for the solutions to differ.

Theorem 5 *Suppose that government utility is given by (49) and that Assumptions 1', 2, 3, 4 and 5 hold. Then for any combination of strategy profiles* Γ *and* $\underline{\alpha}$ *that support a sustainable mechanism, there exists another pair of equilibrium strategy profiles* Γ^* *and* $\underline{\alpha}^* = (\alpha^* \mid \alpha')$ *for some* α' *such that* Γ^* *induces direct submechanisms,* $\underline{\alpha}^*$ *induces truth telling along the equilibrium path, and* $\underline{c}[\Gamma, \underline{\alpha}, \underline{h}] = \underline{c}[\Gamma^*, \underline{\alpha}^*, \underline{h}]$, $l[\Gamma, \underline{\alpha}, \underline{h}] = l[\Gamma^*, \underline{\alpha}^*, \underline{h}]$ *and* $x[\Gamma, \underline{\alpha}, \underline{h}] = x[\Gamma^*, \underline{\alpha}^*, \underline{h}]$. *Moreover, the best sustainable mechanism is a solution to maximizing (9) subject to (10), (11) and the government sustainability constraint:*

$$\mathbb{E}_t \sum_{s=0}^{\infty} \delta^s \left[(1-a)\, v\left(x_{t+s}\left(h^{t+s}\right)\right) + \right.$$
$$\left. a \left(\mathbb{E}_{t+s} \int \left[u\left(c\left(\theta^{t+s}, h^{t+s}\right)\right) - \chi\left(l\left(\theta^{t+s}, h^{t+s}\right) \mid \theta_{t+s}\right)\right] dG^{t+s}\left(\theta^{t+s}\right) \right) \right] \geq$$
$$\max_{\tilde{x}_t' + \int \tilde{c}_t'\left(\theta^t, h^t\right) dG^t\left(\theta^t\right) \leq F\left(K_t\left(h^{t-1}\right), L_t\left(h^t\right)\right)} (1-a)\, v\left(\tilde{x}_t\right) + a \int u\left(\hat{c}_t\left(\theta^t, h^t\right)\right) dG^t\left(\theta^t\right), \quad (50)$$

for all t *and fall* $h^t \in H^t$.

Proof. See Appendix E. ■

The idea of this theorem is that the same type of punishment strategies that were used in the case of the purely self-interested government also work when the government is benevolent. In particular imagine that the government has undertaken a deviation in which it has used some of its past information in order to improve the ex post allocation of resources. This could clearly be desirable given the utility function of the government in (49), but as illustrated with the Roberts' (1984) example, it may have very negative consequences ex ante. Therefore, the best sustainable mechanism will have to discourage such deviations. To do this, imagine the same punishment strategies as above, in which following any type of deviation, all individuals supply zero labor. To establish Theorem 5, all we need to show is that such punishment strategies are sequentially rational. When all other agents choose zero labor supply, following any deviation to positive labor supply, the government would consume some of the increase in output itself, and would redistribute the rest equally among all agents given the separable utility function assumed in Assumption 1'. Since there is a very large number of citizens, this implies the deviating individual will receive no additional consumption from supplying positive labor, and thus it is sequentially rational for all citizens to supply zero labor following a deviation by the government.

This theorem therefore shows that revelation principle applies to the case of benevolent, but time-inconsistent governments as well, though under the additional assumption of Assumption 1'. The next example shows why this assumption is necessary:

Example 1 To avoid issues of deviation among continuum of agents, let us consider a finite economy with n agents for this example, where n is large (exactly the same example can be constructed in an economy with a continuum of agents). There are two types of agents, $\theta \in \{0,1\}$, with $\theta = 0$ corresponding to the disabled type, who can only supply $l = 0$, and has utility $u(c, \cdot \mid \theta = 0) = u(c)$, while the utility of type $\theta = 1$ is $u(c, l \mid \theta = 1) = u(c - \chi_1(l))$, where with $\chi_1(\cdot)$ strictly increasing in l. Furthermore, suppose that aggregate output is linear in labor and that the government is fully benevolent, i.e., $a = 1$ in terms of the utility function in (49). Now imagine the economy has entered the punishment phase where each citizen is supposed to supply $l = 0$ and consume $c = 0$. Consider a deviation by an agent, i', of type $\theta = 1$ to $l' > 0$ such that $\chi_1(l') < 1$. Following this deviation, the benevolent planner will distribute consumption (output $l' > 0$) to maximize its own utility, which involves maximizing average utility of the citizens, thus equating the marginal utility of consumption across agents, i.e.,

$$u'(c_i) = u'\left(c_{i'} - \chi_1\left(l'\right)\right) \text{ for all } i \neq i'$$

thus, $c_{i'} = c_i + \chi_1(l')$ for all $i \neq i'$. The resource constraint is $(n-1)c_i + c_{i'} = l'$, or $c_i = (l' - \chi_1(l'))/n$ and $c_{i'} = (l' - \chi_1(l'))/n + \chi_1(l')$. The resulting utility of individual i' is

$$u\left(\left(l' - \chi_1\left(l'\right)\right)/n\right) > u(0),$$

for any n, thus giving him greater utility than supplying zero labor. This proves that the punishment phase where each citizen is supposed to supply zero labor is not sequentially rational and thus cannot be part of a (Perfect Bayesian) equilibrium with this utility function.

The next theorem provides a generalization of Theorem 4 for the most-commonly studied case where types are constant (in our context, types are "almost constant" as in subsection 6.2) and $\beta = \delta$.[46]

Theorem 6 *Suppose that government utility is given by (49) with $a \in (0,1)$ and that Assumptions 1', 2, 3, 4 and 5 hold. Furthermore, assume that there are (almost) constant types, $\beta = \delta$ and $au'(0) \neq (1-a)v'(0)$. Then, asymptotically there are no aggregate distortions on labor supply and capital accumulation.*

Proof. See Appendix E. ∎

This theorem implies that in an economy with (almost) constant types, aggregate distortions disappear irrespective of the degree of benevolence of the government. Consequently,

[46] Once again with the game form in Remark 1, this theorem can be stated for constant types.

there will be no aggregate capital taxes and no further taxes on labor beyond those implied by the full-commitment Mirrlees economy.[47] In the case where $a \to 1$, the government is arbitrarily close to the fully-benevolent case, and the theorem contrasts with the results in Roberts (1984), where in a very similar environment, the equilibrium always involved extreme distortions. Once again, the main source of the difference is the infinite-horizon nature of our economy, which allows us to construct equilibria in which the government will be punished if it exploits the information it gathers via the earlier submechanisms.

8 Anonymous Markets versus Mechanisms

We have so far characterized the behavior of the best sustainable mechanism under political economy constraints. Although this was largely motivated by our objective of understanding the form of optimal policy in an environment with both informational problems on the side of agents and selfish behavior on the side of the government, an additional motivation is to investigate when certain activities should be regulated by (sustainable) mechanisms and when they should be organized in anonymous markets. In this section, we begin this analysis. Space restrictions preclude a detailed discussion of how anonymous markets should be modeled, so we take the simplest conception of anonymous markets as one in which there is no intervention by the government, and consequently more limited insurance. For the purposes of the exercise in this section, we do not need to assume anything specific about how the anonymous markets work, except that there exists a well-defined anonymous market equilibrium, which yields ex ante utility $\mathbf{U}^{AM}$ to individuals before they know anything about their types. The point to note is that $\mathbf{U}^{AM}$ is independent of both the discount factor of the government and any other institutional controls imposed on the government (since there is no government involvement in the anonymous markets).

Given this, we can provide some simple comparisons between anonymous markets versus sustainable mechanisms. Throughout this section, we suppress dependence of strategies and allocations on public histories to simplify notation. Our first comparative static result states that an increase in the discount factor of the government, δ, makes mechanisms more attractive relative to markets. Let $\mathbf{U}^{SM}(\delta)$ be the ex ante expected value of the best sustainable mechanism when the government discount factor is δ and $\mathbf{U}^{AM}$ be as defined above.

Proposition 3 *Suppose* $\mathbf{U}^{SM}(\delta) \geq \mathbf{U}^{AM}$, *then* $\mathbf{U}^{SM}(\delta') \geq \mathbf{U}^{AM}$ *for all* $\delta' \geq \delta$. *Moreover, as* $\delta \to 0$, $\mathbf{U}^{AM} > \mathbf{U}^{SM}(\delta)$.

[47] The assumption that $au'(0) \neq (1-a)v'(0)$ rules out a special case in which our method of proof does not work (though other more complicated approaches may work even without this assumption).

Proof. Let $\mathcal{S}(\delta)$ be the feasible set of allocation rules when the government discount factor is equal to δ (meaning that they are feasible and also satisfy the sustainability constraint (12)). Let $\{c_t(\delta), l_t(\delta), x_t(\delta)\}_{t=0}^{\infty} \in \mathcal{S}(\delta)$ represent the best sustainable mechanism, where $c_t(\delta)$ and $l_t(\delta)$ are vectors consumption and labor supply across types. Since $\delta' \geq \delta$, we immediately have $\{c_t(\delta), l_t(\delta), x_t(\delta)\}_{t=0}^{\infty} \in \mathcal{S}(\delta')$—when the government's discount factor is δ', the left-hand side of (12) is higher, while the right-hand side is unchanged, so $\{c_t(\delta), l_t(\delta), x_t(\delta)\}_{t=0}^{\infty}$ satisfies (12). Therefore, $\{c_t(\delta), l_t(\delta), x_t(\delta)\}_{t=0}^{\infty}$ is feasible and yields expected utility $\mathbf{U}^{SM}(\delta)$ when the government's discount factor is δ'. This implies that $\mathbf{U}^{SM}(\delta')$ is at least as large as $\mathbf{U}^{AM}$, therefore $\mathbf{U}^{SM}(\delta') \geq \mathbf{U}^{SM}(\delta) \geq \mathbf{U}^{AM}$.

The second part follows from the observation that with anonymous markets, individuals can always achieve the autarchy allocation, thus $\mathbf{U}^{AM} \geq \mathbb{E}_0\left[\sum_{t=0}^{\infty} \beta^t u(c^a(\theta_t), l^a(\theta_t) \mid \theta_t)\right]$,where c^a and l^a denote the optimal autarchy choices of an agent with type θ. In contrast, with $\delta \to 0$, the centralized mechanism leads to utility of $\mathbb{E}_0\left[\sum_{t=0}^{\infty} \beta^t u(0, 0 \mid \theta_t)\right] < \mathbb{E}_0\left[\sum_{t=0}^{\infty} \beta^t u(c^a(\theta_t), l^a(\theta_t) \mid \theta_t)\right]$.
■

An important implication of this proposition is that because the government is self-interested and unable to commit to policy sequences, not all equilibrium allocations without government intervention can be achieved by a mechanism operated by the government. Consequently, as shown by the last part of Proposition 3, anonymous markets can be preferred to sustainable mechanisms.

Let us next consider a modification of our main setup along the lines mentioned in footnote 17, whereby the government can consume only a portion of the output η, so that a lower η corresponds to better insututional constraints on government behavior. Define the value of the mechanism as $\mathbf{U}^{SM}(\eta)$, i.e., now as a function of the institutional restriction on the government. We then have:

Proposition 4 *Suppose* $\mathbf{U}^{SM}(\eta) \geq \mathbf{U}^{AM}$, *then* $\mathbf{U}^{SM}(\eta') \geq \mathbf{U}^{AM}$ *for all* $\eta' \leq \eta$. *Moreover, as* $\eta \to 0$, $\mathbf{U}^{SM}(\eta) > \mathbf{U}^{AM}$.

The proof of this proposition is similar to that of Proposition 3 and is omitted. It states the intuitive result that better institutional controls on government make mechanisms more desirable relative to markets. It also implies that with sufficiently good controls on centralized mechanisms (government behavior), sustainable mechanisms are preferred to anonymous markets.

9 Concluding Remarks

The optimal taxation literature pioneered by Mirrlees (1971) has generated a number of important insights about the optimal tax policy in the presence of insurance-incentive trade-offs. The recent optimal dynamic taxation literature has extended these insights to a macroeconomic setting where issues of dynamic behavior of taxes is of central importance. A potential criticism against all of this literature is that these optimal tax schemes assume a benevolent government with full commitment power. A relevant and important question in this context is whether the insights of this literature apply to real world situations where politicians care about reelection, self-enrichment or their own individual biases, and cannot commit to sequences of future policies or to mechanisms.

This paper investigated this question and characterized the conditions under which these insights hold even when mechanisms are operated by self-interested politicians, who can misuse the resources and the information they collect. The potential misuse of resources and information by the government (politicians or bureaucrats) makes mechanisms less desirable relative to markets than in the standard mechanism design approach, and implies that certain allocations resulting from anonymous market transactions will not be achievable via centralized mechanisms. Nevertheless, centralized mechanisms may be preferable to anonymous markets because of the additional insurance they provide to risk-averse agents.

The main contribution of the paper is an analysis of the form of mechanisms to insure idiosyncratic (productivity) risks as in the classical Mirrlees setup in the presence of a self-interested government. Given the infinite horizon nature of the environment in question, we can construct *sustainable mechanisms* where the government is given incentives not to misuse resources and information. An important result of our analysis is the *revelation principle along the equilibrium path*, which shows that truth-telling mechanisms can be used despite the commitment problems and the different interests of the government and the citizens. Using this tool, we provide a characterization of the best sustainable mechanism.

The other major results of our analysis are as follows. First, under fairly general conditions, the best sustainable mechanism is a solution to a quasi-Mirrlees problem, defined as a program in which the ex ante utility of agents is maximized subject to incentive compatibility, feasibility constraints as well as two additional constraints on the total amount of consumption and labor supply in the economy. Second, under additional conditions, we can characterize the initial and asymptotic distortions created by the best sustainable mechanism. In particular, when the government is sufficiently patient (in many cases as patient as, or more patients than, the citizens), we can show that the Lagrange multiplier on the sustainability constraint of the

government goes to zero and aggregate distortions disappear asymptotically. Consequently, in the long run the highest type individuals will face zero marginal tax rate on their labor supply as in classical Mirrlees setup and there will be no aggregate capital taxes as in the classical dynamic taxation literature. These latter results therefore imply that some of the insights from Mirrlees' classical analysis and from the dynamic taxation literature may follow despite the presence of political economy constraints and commitment problems. However, we also show that when the government is not sufficiently patient, aggregate distortions remain, even asymptotically. In this case, in contrast to many existing studies of optimal taxation, there will be positive distortions and positive aggregate capital taxes even in the long run.

In addition, assuming that individual preferences are separable between consumption and leisure, we also generalized the results on revelation principle and direct mechanisms to an environment with potentially benevolent preferences for the government (including the fully-benevolent time-inconsistent case), and showed that our main characterization result applies with (almost) constant types.

We view this paper as a first step in investigating political economy of mechanisms. There are both technical and substantial issues left unanswered. First, we would like to generalize the results on time-inconsistent fully-benevolent government to non-separable utility functions and to richer dynamics of individual types. Secondly, it is important to undertake a more detailed comparison of centralized mechanisms subject to commitment problems and government misbehavior to more realistic models of anonymous markets. Finally and perhaps most importantly, our investigation highlights a difficult but interesting question: how should the society be structured so that the government (the mechanism designer) is easier to control. In other words, the recognition that governments need to be given the right incentives in designing mechanisms opens the way for the analysis of "mechanism design squared", where the structure of incentives and institutions for governments and individuals are simultaneously determined (for example, in the form of "constitutional design"). This becomes relevant in particular when we want to think about the interaction of different types of institutions in society, for example the difference between contracting institutions that regulate the relationship between individual citizens versus "property rights institutions" that regulate the relationship between the state and individuals (Acemoglu and Johnson, 2005). While the existence of these distinct types of institutions have been recognized, how they should be simultaneously designed has not been investigated. We believe that the approach and tools in this paper will be useful to address this class of questions.

Appendices

10 Appendix A: Some Technical Results and Definitions

We now provide a number of technical results and definitions that will be used in the rest of the appendix. We take the definition of regular point from Luenberger (1969, p. 240).

Definition 6 *Let X and Z be Banach spaces and $G : X \to Z$ be a vector-valued mapping. Suppose that G is continuously (Fréchet) differentiable in the neighborhood of x_0 with the derivative denoted by $G'(x_0)$. Then x_0 is said to be a regular point of G if $G'(x_0)$ maps X onto Z.*

Lemma 2 *Let X and Z be Banach spaces. Consider the maximization problem of*

$$P(u) = \max_{x \in X} f(x) \tag{51}$$

subject to

$$g_0(x) \leq u \tag{52}$$

and

$$G(x) \leq \mathbf{0} \tag{53}$$

where $f : X \to \mathbb{R}$ and $g_0 : X \to \mathbb{R}$ are real-valued functions and $G : X \to Z$ is a vector-valued mapping and $\mathbf{0}$ is the zero of the Banach space Z. Suppose that f is concave and g_0 is convex, and moreover that the solution at $u = 0$, x_0, is a regular point. Let μ be any multiplier of (52). Then μ is a subgradient of $P(0)$.

Proof. This lemma is a direct generalization of Proposition 6.5.8 of Bertsekas, Nedic and Ozdaglar (2003, p. 382) to an infinite dimensional maximization problem. ■

Theorem 7 *Let X and Z be Banach spaces. Consider the maximization problem of*

$$P(\mathbf{u}) = \max_{x \in X} f(x)$$

subject to

$$G(x) \leq \mathbf{0} + \mathbf{u}$$

where $f : X \to \mathbb{R}$ is a real-valued concave function and and $G : X \to Z$ is a convex vector-valued mapping and $\mathbf{0}$ is the zero of the vector space Z and $\mathbf{u}$ is a perturbation. Suppose that x_0 is a solution to this program. Suppose also that x_0 is a regular point of G and that f and G are continuously (Fréchet) differentiable in the neighborhood of x_0. Then $P(\mathbf{0})$ is differentiable.

Proof. From Lemma 2, it follows that if there is a unique multiplier, P has a unique subgredient and is thus differentiable. Proposition 4.47 in Bonnans and Shapiro (2000) establishes that under a weaker constraint qualification condition than regularity, this problem has a unique multiplier. ■

Next, consider a metric space X and $\mathbf{M}(X)$ to be the space of all measures defined on Borel sets of X. An element $\xi \in \mathbf{M}(X)$ is nonnegative, countably additive, and has the property that $\xi(X) = 1$. Let $\mathbf{C}(X)$ be the space of all bounded real-valued continuous functions on X. Then, following Parthasarathy (1967, p. 40) we have that:

Definition 7 *The weak topology on $\mathbf{M}(X)$ is the topology with the following sets as basis:*

$$\begin{aligned} \textit{For any } \xi &\in \mathbf{M}(X),\ f_1, ..., f_k \in \mathbf{C}(X) \textit{ and } \varepsilon_1, ..., \varepsilon_k \in \mathbb{R}_+, \\ S_\xi(f_1, ..., f_k; \varepsilon_1, ..., \varepsilon_k) &= \left\{ \nu : \nu \in \mathbf{M}(X), \left| \int f_i d\xi - f_i d\nu \right| < \varepsilon_i \textit{ for } i = 1, ..., k \right\}. \end{aligned}$$

The key result for us is Theorem 6.5 from Parthasarathy (1967, p. 45):

Theorem 8 *Let X be a compact metric space, then $\mathbf{M}(X)$ is a compact metric space with the weak topology.*

Next, recall the Caratheodory's Theorem, with $conv(X)$ denoting the convex hull of X for some $X \in \mathbb{R}^N$ (e.g., Proposition 1.3.1 in Bertsekas, Nedic and Ozdaglar, 2003, pp. 37-38):

Theorem 9 *(Caratheodory's Theorem) Let $X \in \mathbb{R}^N$, then any $x \in conv(X)$ can be represented as a convex combination of vectors $x_1, ..., x_m$ from X such that $x_2 - x_1, ..., x_m - x_1$ are linearly independent.*

Corollary 1 *Let $X \in \mathbb{R}^N$, then any $x \in conv(X)$ may be represented by no more than $N + 1$ vectors of X.*

Proof. Suppose not, this would violate linear independence of $x_2 - x_1, ..., x_m - x_1$ as stated in Theorem 9 ■

This theorem and its corollary imply that the convex hull of any subset of the N-dimensional Euclidean space can be achieved by $N + 1$ vectors, and will be useful in reducing the dimension of randomizations below.

11 Appendix B: Proofs for Section 3

Proof of Lemma 1: If (8) is violated following some public history h^t, then $\xi_t(h^t) = 1$, $\tilde{x}'_t(h^t) = F(K_t(h^{t-1}), L_t(h^t))$ and $\tilde{c}'_t = \tilde{c}'^{\emptyset}_t$ yields utility $v(F(K_t(h^{t-1}), L_t(h^t)))$ to the government, which is greater than its equilibrium payoff following h^t, given by the left-hand side of (8). This contradicts sustainability and establishes that (8) is necessary in any sustainable mechanism.

To see that (8) is sufficient for the best sustainable mechanism, note that reducing $v^c_t\left(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t\right)$ is equivalent to relaxing the constraint on problem (4), so is always preferred. Since from Assumption 5, $v^c_t \geq 0$ (i.e., $x \geq 0$ and $v(0) = 0$), we only need to show that $v^c_t\left(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t\right) = 0$ is achievable for all $\tilde{M}^t \in \mathcal{M}^t$, $\Gamma' \in \mathcal{G}$, $\tilde{K}'_{t+1} \in \mathbb{R}_+$ and $\tilde{c}'_t \in \mathcal{C}_t$. The following simple combination of strategies would achieve this objective. Let ρ^t be the history of actions by the government. Also denote $\tilde{c}'_t = \tilde{c}'^{\emptyset}_t$ be the mapping that allocates zero consumption to all individuals irrespective of past and current reports. Let $\rho^t = \hat{\rho}^t$ if $\tilde{x}_{t-s}(h^{t-s}) = x_{t-s}(h^{t-s})$ and $\tilde{M}_{t-s} = M_{t-s}$ for all $s > 0$. Then the following strategy combination would ensure $v^c_t\left(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t\right) = 0$ for all t: (1) for the citizens, $\underline{\alpha} = (\tilde{\alpha} \mid \alpha^{\emptyset})$, for some $\tilde{\alpha}$, which means that for each citizen i and for all t, we have that if $\rho^{t-1} = \hat{\rho}^{t-1}$, then $\alpha^i_t = \tilde{\alpha}$, and if $\rho^{t-1} \neq \hat{\rho}^{t-1}$, then $\alpha^i_t = \alpha^{\emptyset}$; (2) for the government, Γ, such that if $\rho^{t-1} = \hat{\rho}^{t-1}$, then Γ involves $\tilde{x}_t(h^t) = x_t(h^t)$, $\tilde{M}_t = M_t$, and $\xi_t(h^t) = 0$ for all $h^t \in H^t$; and if $\rho^{t-1} \neq \hat{\rho}^{t-1}$, then it involves $\xi_t(h^t) = 1$, $\tilde{x}'_t(h^t) = F(K_t(h^{t-1}), L_t(h^t))$ for all $h^t \in H^t$, and $\tilde{c}'_t = \tilde{c}'^{\emptyset}_t$.

We next need to show that these strategies are sequentially rational. Consider the citizens first; it suffices to note that following a history where $\rho^{t-1} \neq \hat{\rho}^{t-1}$, the government is playing $\xi_{t+s}(h^{t+s}) = 1$, $\tilde{x}'_{t+s}(h^{t+s}) = F(K_{t+s}(h^{t+s-1}), L_{t+s}(h^{t+s}))$ and $\tilde{c}'_{t+s} = \tilde{c}'^{\emptyset}_{t+s}$for all $s \geq 0$. Therefore, any strategy other than $\alpha^{\emptyset}$ will give some utility less than $\mathbb{E}\left[\sum_{s=0}^{\infty} \beta^s u(0, 0 \mid \theta_{t+s}) \mid \theta^t, h^t\right]$ to an individual with type history θ^t in the current period, which is the utility that always playing $\alpha^{\emptyset}$ delivers. This argument proves that this strategy is sequentially rational for the citizens. It is also sequentially rational for the government, since after any history of $\rho^{t-1} \neq \hat{\rho}^{t-1}$, there will be no future output to expropriate, thus playing $\xi_{t+s}(h^{t+s}) = 1$, $\tilde{x}'_{t+s}(h^{t+s}) = F(K_{t+s}(h^{t+s-1}), L_{t+s}(h^{t+s}))$ and $\tilde{c}'_{t+s} = \tilde{c}'^{\emptyset}_{t+s}$ is a best response for the government starting in all of its information sets for all $s \geq 0$. The fact that $v^c_t\left(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t\right) = 0$ for all $\tilde{M}^t \in \mathcal{M}^t$, $\tilde{K}_{t+1} \in \mathbb{R}_+$ and $\tilde{c}'_t \in \mathcal{C}_t$, then implies that the best deviation for the government is

also $\xi_t(h^t) = 1$, $\tilde{x}'_t(h^t) = F\left(K_t\left(h^{t-1}\right), L_t\left(h^t\right)\right)$, $\tilde{c}'_t = \tilde{c}'^{\emptyset}_t$ and $\tilde{K}'_{t+1}(h^t) = 0$ for all $h^t \in H^t$ establishing that (7) takes the form (8).■

Proof of Theorem 1: Take equilibrium strategy profiles Γ and $\underline{\alpha}$ that support a sustainable mechanism. Then by definition $\xi_t(h^t) = 0$ for all t and $h^t \in H^t$, and from Lemma 1, (8) is satisfied. Let the best response of type θ^t at time t according to $\underline{\alpha}$ be to announce $z_{t,\Gamma}\left(\theta^t, h^{t-1}\right)$ given a history of reports $z_\Gamma^{t-1}\left(\theta^{t-1}, h^{t-1}\right)$ and public history h^{t-1}. Let
$z_\Gamma^t\left(\theta^t, h^{t-1}\right) = \left(z_\Gamma^{t-1}\left(\theta^{t-1}, h^{t-1}\right), z_{t,\Gamma}\left(\theta^t, h^t\right)\right)$. Denote the expected utility of this individual under this mechanism given history h^{t-1} be $\tilde{u}\left[z_\Gamma^t\left(\theta^t, h^{t-1}\right) \mid \theta^t, \Gamma, h^{t-1}\right]$. By definition of $z_\Gamma^t\left(\theta^t, h^{t-1}\right)$ being a best response, we have

$$\tilde{u}\left[z_\Gamma^t\left(\theta^t, h^{t-1}\right) \mid \theta^t, \Gamma, h^{t-1}\right] \geq \tilde{u}\left[\tilde{z}_\Gamma^t\left(\theta^t, h^{t-1}\right) \mid \theta^t, \Gamma, h^{t-1}\right] \text{ for all } \tilde{z}_\Gamma^t\left(\theta^t, h^{t-1}\right) \in Z^t \text{ and } h^{t-1} \in H^{t-1}.$$

Now consider the alternative strategy profile for the government Γ^*, which induces the action profile $\left[\left\{\tilde{M}_t, \xi_t(h^t), \tilde{x}_t(h^t), \tilde{x}'_t(h^t), \tilde{c}'_t\right\}_{t=0}^{\infty}\right]$ such that $\xi_t(h^t) = 0$ for all t and $h^t \in H^t$, $\tilde{M}_t = M_t^*$ (where M_t^* is a direct submechanism) and $\underline{c}[\Gamma^*, \underline{\alpha}^*, \underline{h}] = \underline{c}[\Gamma, \underline{\alpha}, \underline{h}]$, $\underline{l}[\Gamma^*, \underline{\alpha}^*, \underline{h}] = \underline{l}[\Gamma, \underline{\alpha}, \underline{h}]$, and $x[\Gamma, \underline{\alpha}, \underline{h}] = x[\Gamma^*, \underline{\alpha}^*, \underline{h}]$. Therefore, by construction,

$$\begin{aligned} \tilde{u}\left[\theta^t, h^{t-1} \mid \theta^t, \Gamma^*, h^{t-1}\right] &= \tilde{u}\left[z_\Gamma^t\left(\theta^t, h^{t-1}\right) \mid \theta^t, \Gamma, h^{t-1}\right] \\ &\geq \tilde{u}\left[\tilde{z}_\Gamma^t\left(\theta^t, h^{t-1}\right) \mid \theta^t, \Gamma, h^{t-1}\right] = \tilde{u}\left[\hat{\theta}^t, h^{t-1} \mid \theta^t, \Gamma^*, h^{t-1}\right] \end{aligned} \quad (54)$$

for all $\hat{\theta}^t \in \Theta^t$ and all $h^{t-1} \in H^{t-1}$. Equation (54) implies that $\underline{\alpha}^* = (\alpha^* \mid \alpha')$ is a best response along the equilibrium path for the agents against the mechanism M^* and government strategy profile Γ^*. Moreover, by construction, the resulting allocation when individuals play $\underline{\alpha}^* = (\alpha^* \mid \alpha')$ against Γ^* is the same as when they play $\underline{\alpha}$ against Γ. Therefore, by the definition of Γ being sustainable, we have $\Gamma \succeq_{\underline{\alpha}} \Gamma'$ for all $\Gamma' \in \mathcal{G}$. Now choose α' to be identical to $\underline{\alpha}$ off-the-equilibrium path, which implies that $\Gamma^* \succeq_{\underline{\alpha}^*} \Gamma'$ for all $\Gamma' \in \mathcal{G}$ or that (8) is satisfied, thus establishing that $(\Gamma^*, \underline{\alpha}^*)$ is an equilibrium.■

12 Appendix C: Proofs for Section 5

In this appendix, we provide and prove a number of results used in the analysis of Section 5, ultimate the building up to the proof of Theorem 3. Throughout this section we assume that Assumptions 1-3 and 5 are satisfied.

12.1 Properties of the Function $U(C, L)$

As mentioned in the text, to establish the concavity of $U(C,L)$, we follow Prescott and Townsend (1984a, 1984b) and allow for stochastic mechanisms, i.e., $M_t \equiv \left(\tilde{c}_t, \tilde{l}_t\right) : Z^t \times H^t \to \Delta\left(\mathbb{R}_+ \times [0, \bar{l}]\right)$ as specified in footnote 16. Recall that $U(C,L)$ is a solution to a finite-dimensional maximization problem. Moreover, using the single-crossing property (Assumption 2), we can reduce the static incentive compatibility constraints to only the constraints for the neighboring types, thus to N constraints (there is no incentive compatibility constraint for the lowest type, θ_0). In addition, there are the resource constraints on the sum of consumption and labor supply levels. Recall also that only $(C,L) \in \Lambda$ will enable this maximization program to be well defined by making the constraint set non-empty.

Let $\mathcal{C} = \{(c,l) \in \mathbb{R}^2 : 0 \leq c \leq \bar{c},\ 0 \leq l \leq \bar{l}\}$ be the set of possible consumption-labor allocations for agents. Let $\mathcal{P}$ be the space of $N+1$-tuples of probability measures on Borel subsets of $\mathcal{C}$. Thus each element $\zeta = [\zeta(\theta_0), ..., \zeta(\theta_N)]$ in $\mathcal{P}$ consists of $N+1$ probability measures for each type $\theta_i \in \Theta$. Let us also denote the fraction of individuals with type θ at any point in time by $\pi(\theta)$, where clearly $\sum_{i=0}^{N} \pi(\theta_i) = 1$.

Then the quasi-Mirrlees problem can be defined in the following way

$$U(C\left(h^{t}\right), L\left(h^{t}\right)) \equiv \max_{\zeta(\cdot|h^{t})\in\mathcal{P}} \sum_{\theta\in\Theta} \pi(\theta) \int u(c, l; \theta)\zeta(d(c, l), \theta \mid h^{t}) \tag{55}$$

subject to

$$\int u(c, l \mid \theta_{i})\zeta(d(c, l), \theta_{i} \mid h^{t}) \geq \int u(c, l \mid \theta)\zeta(d(c, l), \theta_{i-1} \mid h^{t}) \text{ for all } i = 1, ..., N \tag{56}$$

$$\sum_{\theta\in\Theta} \pi(\theta) \int c\zeta(d(c, l), \theta \mid h^{t}) \leq C\left(h^{t}\right) \tag{57}$$

$$\sum_{\theta\in\Theta} \pi(\theta) \int l\zeta(d(c, l), \theta \mid h^{t}) \geq L\left(h^{t}\right) \tag{58}$$

for some $(C, L) \in \Lambda$.

Before deriving properties of the function $U(C, L)$, we need to ensure regularity. Let (56), (57) and (58) define the constraint mapping.

Lemma 3 *The solution to (55) is a regular point of the constraint mapping.*

Proof. The proof follows from the fact that from single-crossing property (Assumption 2), all incentive compatibility constraints in (56) are linearly independent from each other, and also linearly independent from (57) and (58), thus the constraint mapping has full rank, $N+2$, and is thus onto. ■

Our main result on the function $U(C, L)$ is:

Lemma 4 *$U(C, L)$ is well-defined, continuous and concave on Λ, nondecreasing in C and nonincreasing in L and differentiable in (C, L).*

Proof. First, we show that $U(C, L)$ is well-defined, i.e., a solution exists. For this, endow the set of probability measures $\mathcal{P}$ with the weak topology. Since $\mathcal{C}$ is a compact subset of $\mathbb{R}^2$, Theorem 8 above establishes that $\mathcal{P}$ is compact in the weak topology, and the constraint set is compact in the weak topology as well. Moreover, the objective function is continuous in any $\zeta \in \mathcal{P}$, thus establishing existence.

Next, to show that $U(C, L)$ is continuous, note that with the lotteries the constraint set is convex. Then from Berge's Maximum Theorem (e.g., Stokey, Lucas and Prescott, 1989, Theorem 3.6, p. 62), $U(C, L)$ is continuous in (C, L).

Concavity then follows from the convexity of the constraint set and the fact that the objective function is concave in $\zeta \in \mathcal{P}$.

$U(C, L)$ is also clearly nondecreasing in C, since a higher C relaxes constraint (57) and nonincreasing in L, since a higher L tightens constraint (58).

Finally, to prove differentiability, note that the regularity condition is satisfied from Lemma 3 and moreover, the objective function in (55) is continuously differentiable at all points of the constraint set. We can therefore apply Theorem 7 using the strong topology, establishing that $U(C, L)$ is differentiable in (C, L). This completes the proof of the lemma. ■

The necessary properties of the set Λ are derived in the next lemma.

Lemma 5 *Λ is compact and convex.*

Proof. Convexity: Consider $(C^0, L^0), (C^1, L^1) \in \Lambda$ and some ζ^0, ζ^1 feasible for (C^0, L^0) and (C^1, L^1) respectively. For any $\alpha \in (0, 1)$ $\zeta^{\alpha} \equiv \alpha\zeta^0 + (1-\alpha)\zeta^1$ is feasible for $(\alpha C^0 + (1-\alpha)C^1, \alpha L^0 + (1-\alpha)L^1)$, so that this set is non-empty. Moreover, since ζ^0, ζ^1 satisfy (56), (57) and (58), ζ^{α} satisfies all three of these constraints, establishing convexity.

Compactness: Λ is clearly bounded, so we only have to show that it is closed. Take a sequence $(C^n, L^n) \in \Lambda$. Since this sequence is in a bounded set, it has a convergent subsequence, $(C^n, L^n) \to (C^\infty, L^\infty)$. We just need to show that $(C^\infty, L^\infty) \in \Lambda$. Let ζ^n be a feasible element for (C^n, L^n), and since $\mathcal{P}$ is compact under the weak topology, $\zeta^n \to \zeta^\infty \in \mathcal{P}$, which implies that ζ^∞ satisfies (56)-(58) and so ζ^∞ is feasible for (C^∞, L^∞), therefore Λ is closed. ∎

Now define a promised utility for the government for some sequence $x = \{x_t\}_{t=0}^\infty$ as

$$w = \sum_{t=0}^{\infty} \delta^t v(x_t)$$

Then the set of feasible promised utilities $\mathbb{W}$ is defined as

$$\mathbb{W} = \{w : \exists x \in \mathbb{R}^\infty \text{ s.t. for any } t \text{ there is some } L \text{ s.t. } (L - x_t, L) \in \Lambda,\ w = \sum_{t=0}^{\infty} \delta^t v(x_t)\}$$

Lemma 6 $\mathbb{W} = [0, \bar{w}]$.

Proof. Since $v(0) = 0$, it is clear that 0 is the minimal element. By definition $\bar{w}$ is the maximal element. Moreover, clearly any $w \leq \bar{w}$ is also achievable, so $\mathbb{W}$ must take the form $[0, \bar{w}]$. ∎

To further analyze the best sustainable mechanism, let us rewrite the maximization problem recursively as in equations (27)-(30) in the text. The following lemma is immediate:

Lemma 7 *The solution to (24) subject to (23) and (25) is equivalent to the solution to the program (27)-(30) combined with a choice of initial promised value to the government,* w_0, *such that* $w_0 = \arg\max_{w \in \mathbb{W}} V(w)$.

Proof. The proof follows from Thomas and Worrall (1990). Clearly any solution to (27)-(30) gives a sustainable mechanism. Moreover, the ex ante utility for the citizens from any sustainable mechanism can be obtained as $V(w)$ from (27)-(30) by an argument analogous to the principle of optimality (see, e.g., Stokey, Lucas and Prescott, 1989). It then follows that $V(w_0) = \max_{w \in \mathbb{W}} V(w)$ gives the best sustainable mechanism. ∎

Next note that the constraint set in the program (27)-(30) is not convex, and randomizations over the current consumption and the continuation value of the government may further improve the value of the program (which is the reason why we introduced histories h^t). So analogously to the quasi-Mirrlees problem, we now consider randomizations. Now let $\mathbf{q} = (C, L, x, w') \in \mathbb{R}^4$, $\mathcal{C}(w) = \{\mathbf{q} \in \mathbb{R}^4 : \text{(27)-(30) are satisfied for given } w\}$, and let $\mathcal{Z}$ be the set of Borel subsets of $\mathcal{C}(w)$ Then let the triple $(\mathcal{C}(w), \mathcal{Z}, \bar{\mu})$ be a probability space. Let $\mathcal{P}(w)$ be the space of probability measures on $\mathcal{C}(w)$ endowed with the weak topology. Incorporating randomization, we can write the recursive formulation as:

Problem A1

$$V(w) = \max_{\xi \in \mathcal{P}(w)} \int [U(C, L) + \beta V(w')] \xi(d\mathbf{q}) \tag{59}$$

subject to

$$C + x \leq L \ \xi\text{-almost-surely} \tag{60}$$

$$v(x) + \delta w' \geq v(L) \ \xi\text{-almost-surely} \tag{61}$$

$$w = \int [v(x) + \delta w'] \xi(d\mathbf{q}) \tag{62}$$

and

$$(C, L) \in \Lambda \text{ and } w' \in \mathbb{W} \ \xi\text{-almost-surely.} \tag{63}$$

Note that the resulting solution to this program will correspond to stochastic sequences $\{x_t(h^t)\}_{t=0}^\infty$ and $\{w_t(h^t)\}_{t=0}^\infty$ as specified in the text.

Lemma 8 *$V(w)$ is concave.*

Proof. Consider any w_0 and w_1 and ξ_0 and ξ_1 that are the solution to the maximization problem. Consider $w = \alpha w_0 + (1-\alpha)w_1$ for some $\alpha \in (0,1)$. Let $\xi_\alpha = \alpha\xi_0 + (1-\alpha)\xi_1$. Constraints (60) and (61) hold state by state, and are satisfied for both ξ_0 and ξ_1, and therefore must be satisfied for ξ_α. Constraint (62) is linear in ξ, therefore ξ_α also satisfies this constraint. Since the objective function is linear in ξ_α, we have $V(\alpha w_0 + (1-\alpha)w_1) \geq \alpha V(w_0) + (1-\alpha)V(w_1)$, establishing the concavity of V. ■

The above lemma establishes the concavity of V using arbitrary randomizations in the maximization problem (59). The next lemma shows that a particularly simple form of randomization is sufficient to achieve the maximum of (59).

Lemma 9 *There exists $\xi \in \mathcal{P}(w)$ achieving the value $V(w)$ with randomization between at most two points, (C_0, L_0, x_0, w_0') and (C_1, L_1, x_1, w_1') with probabilities ξ_0 and $1-\xi_0$.*

Proof. To achieve convexity, we only need the constraint set to be convex. The constraint set here is $\mathcal{C}(w) \in \mathbb{R}^4$. Recall Theorem 9 and its corollary, which imply that the convex hull of $\mathcal{C}(w)$ can be achieved with 5 points.

Suppose, to obtain a contradiction, that there are more than two points with positive probability. We consider a case of three points, since the same argument applies to any finite number of points. Suppose that randomization occurs between (C_0, L_0, x_0, w_0'), (C_1, L_1, x_1, w_1') and (C_2, L_2, x_2, w_2') with probabilities $\xi_0, \xi_1, \xi_2 > 0$. Suppose without loss of generality that $v(x_0) + \delta w_0' \leq v(x_2) + \delta w_2' \leq v(x_1) + \delta w_1'$ and let $\alpha \in [0,1]$ be such that $v(x_2) + \delta w_2' = \alpha(v(x_0) + \delta w_0') + (1-\alpha)(v(x_1) + \delta w_1')$. Suppose first

$$U(C_2, L_2) + \beta V(w_2') > \alpha[U(C_0, L_0) + \beta V(w_0')] + (1-\alpha)[U(C_1, L_1) + \beta V(w_1')].$$

Then an alternative element $\hat{\xi} \in \mathcal{P}(w)$ assigning probability $\hat{\xi}_2 = 1$ to (C_2, L_2, x_2, w_2') is feasible and yields higher utility than the original randomization, yielding a contradiction. Next suppose that

$$U(C_2, L_2) + \beta V(w_2') < \alpha[U(C_0, L_0) + \beta V(w_0')] + (1-\alpha)[U(C_1, L_1) + \beta V(w_1')].$$

Now consider an alternative $\hat{\xi} \in \mathcal{P}(w)$ assigning probability $\xi_0 + \alpha\xi_2$ to (C_0, L_0, x_0, w_0') and probability $\xi_1 + (1-\alpha)\xi_2$ to (C_1, L_1, x_1, w_1'), which is again feasible and gives a higher utility than original randomization, once again yielding a contradiction. Therefore, ξ must satisfy

$$U(C_2, L_2) + \beta V(w_2') = [U(C_0, L_0) + \beta V(w_0')] + (1-\alpha)[U(C_1, L_1) + \beta V(w_1')].$$

But then the optimum can be achieved by simply randomizing between (C_0, L_0, x_0, w_0') and (C_1, L_1, x_1, w_1') with probabilities $\xi_0 + \alpha\xi_2$ to (C_0, L_0, x_0, w_0') and probability $\xi_1 + (1-\alpha)\xi_2$. ■

Lemma 9 implies that we can focus on randomizations between two points. We denote solution for any w by $C_i(w), L_i(w), x_i(w), w_i'(w), \xi_i(w)$ for $i \in \{0,1\}$, and rewrite Problem A1 in equivalent form:

Problem A2:

$$V(w) = \max_{\{\xi_i, C_i, L_i, x_i, w_i'\}_{i=0,1}} \sum_{i=0,1} \xi_i \left[U(C_i, L_i) + \beta V(w_i')\right] \tag{64}$$

subject to

$$C_i + x_i \leq L_i \text{ for } i = 0,1 \tag{65}$$

$$v(x_i) + \delta w_i' \geq v(L_i) \text{ for } i = 0,1 \tag{66}$$

$$w = \sum_{i=0,1} \xi_i \left[v(x_i) + \delta w_i'\right]. \tag{67}$$

$$(C_i, L_i) \in \Lambda \text{ for } i = 0,1 \text{ and } w' \in \mathbb{W}. \tag{68}$$

Returning to the notation in the text, the fact that at every date, there is randomization only between two points also implies that the aggregate public history can be taken as $h^t \in \{0,1\}^t$.

Next we would like to establish that $V(w)$ is differentiable. This does not follow from Theorem 7, since $V(w)$ includes the term $V(w_i')$, which may not be differentiable. Instead, we can apply an argument similar to that of Benveniste and Scheinkman (1979) to prove differentiability (see also Stokey, Lucas and Prescott, 1989).

Lemma 10 $V(w)$ *is differentiable.*

Proof. From Lemma 9, in Problem A1 when $w = w_0$, the optimal value can be achieved by randomizing between $\bar{w}_i'(w_0)$ for $i = 0, 1$ with probabilities p_i. Let $V(\bar{w}'(w_0)) = \sum_{i=0,1} p_i V(\bar{w}_i'(w_0))$ and $\bar{w}'(w_0) = \sum_{i=0,1} p_i \bar{w}_i'(w_0))$. Then consider the maximization problem

$$W(w) = \max_{\xi \in \mathcal{P}(w)} \int [U(C,L) + \beta V(\bar{w}'(w_0))] \xi(d\mathbf{q}) \tag{69}$$

subject to (60), (61) and

$$w = \int [v(x) + \delta \bar{w}'(w_0)] \xi(d\mathbf{q}),$$

which only differs from Problem A1 in that $V(w')$ and w' are held constant at $V(\bar{w}'(w_0))$ and $\bar{w}'(w_0)$. By the same argument as in Lemma 8, $W(w)$ is concave (the fact that we have $V(\bar{w}'(w_0))$ fixed in (69) does not affect the proof of Lemma 8). Next the same arguments as in Lemma 9 establishes that $W(w)$ can be equivalently characterized by the following maximization problem:

Problem A3:

$$W(w) = \max_{\{\xi_i, C_i, L_i, x_i\}_{i=0,1}} \sum_{i=0,1} \xi_i [U(C_i, L_i) + \beta V(\bar{w}'(w_0))]$$

$$C_i + x_i \le L_i \text{ for } i = 0, 1 \tag{70}$$

$$v(x_i) + \delta \bar{w}'(w_0) \ge v(L_i) \text{ for } i = 0, 1 \tag{71}$$

they don't

$$w = \sum_{i=0,1} \xi_i [v(x_i) + \delta \bar{w}'(w_0)]. \tag{72}$$

$$(C_i, L_i) \in \Lambda \text{ for } i = 0, 1 \text{ and } w' \in \mathbb{W}. \tag{73}$$

Since $W(w)$ is concave, Theorem 7 implies that it is also differentiable—$\beta V(\bar{w}'(w_0))$ is just a constant here, and all other terms are differentiable. Moreover, we have

$$W(w) \le V(w) \tag{74}$$

and

$$W(w_0) = V(w_0) \tag{75}$$

by construction.

From Lemma 8, $V(w_0)$ is concave, and therefore $-V$ is convex. Convex functions have well-defined subdifferentials. In particular, if f is convex, there exists a closed, convex and nonempty set ∂f such that for all $\nu \in \partial f$ and any x and x', we have $f(x') - f(x) \ge v(x' - x)$ (see Rockafellar, 1970, Chapter 23 or Bertsekas, Nedic and Ozdaglar, 2003, Chapter 4). Let $-\partial V(w)$ be the set of subdifferentials of $-V$, i.e., all $-\nu$ such that $-V(\hat{w}) + V(w) \ge -\nu \cdot (\hat{w} - w)$. By definition, $-\partial V(w)$ is a closed, convex and nonempty set. Consequently, for any subgradient $-\nu$of $-\partial V(w_0)$, we have

$$\nu \cdot (w - w_0) \ge V(w) - V(w_0) \ge W(w) - W(w_0),$$

where the first inequality is by the definition of a subgradient, and the second follows from (74) and (75). This implies that $-\nu$ is also a subgradients of $-W(w_0)$. But since $W(w_0)$ (and thus $-W(w_0)$) is differentiable, $-\nu$ must be unique, therefore $-V(w_0)$ (and thus $V(w_0)$) is also differentiable. ■

12.2 Proof of Theorem 3

Proof. Since V is differentiable from Lemma 4 and concave from Lemma 8, the first-order conditions are necessary and sufficient for the maximization (64). Assign the multipliers $\xi_i\kappa_i$ to the constraints in (65), $\xi_i\psi_i$ to those in (66) and γ to constraint (67), and let $V'(w)$ be the derivative of $V(w)$ at w, we have

$$\begin{aligned}\beta\xi_0 V'(w_0') + \delta\psi_0\xi_0 + \delta\gamma\xi_0 &\leq 0\\ \beta\xi_1 V'(w_1') + \delta\psi_1\xi_1 + \delta\gamma\xi_1 &\leq 0\end{aligned}$$

with both equations holding as equality for $w_i' \in \mathrm{Int}\mathbb{W}$. Therefore,

$$\frac{\beta}{\delta}V'(w_i') \leq -\psi_i - \gamma, \tag{76}$$

again with equality for $w_i' \in \mathrm{Int}\mathbb{W}$. Moreover, since V is differentiable, we have

$$V'(w) \geq -\gamma \tag{77}$$

again with equality for $w \in \mathrm{Int}\mathbb{W}$.

In addition, combining first-order conditions we have that for $(C, L) \in \mathrm{Int}\Lambda$,

$$U_C(C_i, L_i) + U_L(C_i, L_i) = \psi_i v'(L_i) \text{ for } i = 0, 1. \tag{78}$$

Part 1: To establish this part of the theorem, it suffices to show that $\psi_i > 0$ at $t = 0$ for $i = 0$ or 1. First note that the initial value w_0 maximizes $V(w)$, and since $V(\cdot)$ is differentiable, this implies $V'(w_0) = \gamma = 0$ at $t = 0$. Suppose, to obtain a contradiction, that $\psi_i = 0$ at $t = 0$ for both $i = 0$ and 1. This implies from (76) that $\beta V'(w_i')/\delta = 0$, so that $w_i' = w_0$. Repeating this argument yields $w_t = w_0$ for all t, and (66) never binds. This is in turn only consistent with $x_t = 0$ for all t, which then implies $C_t = L_t = 0$ for all t. This cannot be optimal, however, since $\psi_i = 0$ both $i = 0$ and 1 implies $U_C(C_i, L_i) + U_L(C_i, L_i) = 0$, which does not have $C_i = L_i = 0$ as a solution (this follows from Assumption 6, since otherwise $\bar{w}$ would be equal to zero, and the inequality in Assumption 6 could not be strict). This yields a contradiction and establishes that $\psi_i > 0$ for $i = 0$ or 1 at time $t = 0$, so initial (C, L) cannot undistorted, and from Proposition 2, there is a positive marginal tax rate on even the highest type.

Part 2: Let $h^t \in \{0,1\}^t$, $i = 0, 1$ and fix some $w \in \mathrm{Int}\mathbb{W}$. Since $\beta \leq \delta$ and $V'(w) \leq 0$, (76) implies

$$V'(w_i') \leq -\psi_i - \gamma \text{ for } i = 0, 1.$$

Combining this with (77) and $\psi_i \geq 0$ yields:

$$V'(w) \geq V'(w_i') \text{ for } i = 0, 1.$$

Concavity of V then implies that $w_i' \geq w$ for $i = 0, 1$, establishing the claim that the sequence $\{w_t(h^t)\}_{t=0}^{\infty}$ is nondecreasing. Since each $w_t(h^t)$ is in the compact set $[0, \bar{w}]$, the stochastic sequence $\{w_t(h^t)\}_{t=0}^{\infty}$ must converge almost surely to some point w^*, meaning $\operatorname{plim} w_t(h^t) = w^*$. This immediately implies $\{x_t(h^t)\}_{t=0}^{\infty}$ almost surely converges to some x^* (i.e., $\operatorname{plim} x_t(h^t) = x^*$), and also $\operatorname{plim} C_t(h^t) = C^*$ and $\operatorname{plim} L_t(h^t) = L^*$ are feasible (given $\operatorname{plim} x_t(h^t) = x^*$) and are optimal from the concavity of $U(C, L)$, this is a solution to the maximization in (64), establishing the existence of a steady state as claimed in the theorem.

Recall from the above argument that $\{w_t(h^t)\} \uparrow w^*$ almost surely. First suppose that $\beta = \delta$ and $w^* < \bar{w}$. Then we must have $V'(w^*) \leq -\psi_i^* - \gamma^*$ for $i = 0, 1$ and from (77), $V'(w^*) \geq -\gamma^*$, which is only possible if $\psi_i^* = 0$ for $i = 0, 1$ (recall that $\psi_i \geq 0$), establishing the claim that the sustainability constraints, (66), become slack asymptotically. This immediately implies that, asymptotically, the solution to problem (17)-(18) with $\{X_t(h^t)\}_{t=0}^{\infty} \to x^*$ and the solution to problem (19)-(20)-(12) coincide as required for an asymptotically undistorted allocation as specified in Definition 4.

Second, suppose that $\beta = \delta$ and $w^* = \bar{w}$. Recall that $\bar{w} \equiv \max_{(C,L)\in\Lambda} v(L - C)/(1 - \delta)$, and let $(\bar{C}, \bar{L})$ be a solution to this program satisfying the condition in Assumption 6, so that $\bar{x} = \bar{L} - \bar{C}$, and $v(\bar{x})/(1-\delta) > v(\bar{L})$. This implies that at $(\bar{C}, \bar{L}, \bar{x})$, the constraint (12) is slack, and therefore, the solution to problem (19)-(20)-(12) coincides with the solution to problem (17)-(18) with $X(h^t) \to \bar{x}$, so that the best sustainable mechanism is asymptotically undistorted.

Next consider the case in which $\beta < \delta$. Now we have

$$\frac{\beta}{\delta} V'(w_{i,t+1}) + \psi_i \leq V'(w_t) \text{ for } i = 0, 1.$$

Recall that, as established above, $\{w_{i,t}\} \uparrow w^*$. To derive a contradiction, suppose that $w^* < \bar{w}$. This implies that $\beta V'(w^*)/\delta + \psi^* \leq V'(w^*)$ for some $\psi^* \geq 0$, which is impossible in view of the fact that $\beta < \delta$ and $V'(w^*) \leq 0$ (unless $w_t = w_0$ for all t so that $V'(w^*) = 0$, which is ruled out by the argument in part 1). Therefore, we must have $\{w_{i,t}\} \uparrow \bar{w}$. The same argument as in the previous paragraph then establishes that the best sustainable mechanism is asymptotically undistorted.

Part 3: Suppose that $\beta > \delta$. Then, $\{w_t(h^t)\}$ is no longer nondecreasing. If $\{w_t(h^t)\}$ converges to some $w^* \leq \bar{w}$, then equation (34) in the text must hold as $t \to \infty$ and $\lim_{t\to\infty} -U_C/U_L$ exists and is strictly greater than 1 as claimed in the theorem. Next, suppose that $\{w_t\}$ does not converge. Since it lies in a compact set, it has a convergent subsequence. Suppose that for all such convergent subsequences $\psi_i = 0$ for $i = 0, 1$, this would imply convergence to a steady state since we would have $\psi_{i,t} = 0$ for $i = 0, 1$ and for all t, yielding a contradiction. Therefore, there must exist a convergent subsequence with $\psi_i > 0$, so that $\limsup -U_C/U_L > 1$. Consequently, distortions do not disappear asymptotically, completing the proof. ■

13 Appendix D: Proofs for Section 6

In this appendix, we provide the proofs for the more general environment.

13.1 Properties of $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$

As in the above proof, to show concavity and differentiability of $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$, we introduce randomizations. To simplify notation, in this appendix, we suppress dependence on public histories h^t. The original maximization problem without randomization is to maximize (9) subject to (10), (11), and (12) as stated in Proposition 1. Recall also that $\theta_t \in \Theta$, where Θ is at finite set (with $N+1$ elements). Therefore Θ^t for any $t < \infty$ is also a finite set. Consider next the functions $c_t : \Theta^t \to \mathbb{R}_+$ and $l_t : \Theta^t \to [0, \bar{l}]$. By definition, these functions assign values to a finite number of points in the set Θ^t for any $t < \infty$, thus can simply be thought of as vectors of $(N(N+1))^t$ dimension. Moreover .

$$\int c_t(\theta^t)\, dG(\theta^t) \leq \bar{Y},\ K_{t+1} \leq \bar{Y} \text{ and } x_t \leq \bar{Y}, \tag{79}$$

where $\bar{Y} = F(\bar{Y}, \bar{l}) < \infty$. Therefore, $X_t = \{c_t(\theta^t), l(\theta^t), K_{t+1}, x_t\}$ is a vector (of dimension $(N(N+1))^{2t} + 2$). Let $\mathbf{X}_t$ be the set of all such vectors that satisfy the inequalities in (79), and for $X_t \in \mathbf{X}_t$, let $X_t(i)$ denote the ith component of this vector, and T_t be the dimension of vectors in the set $\mathbf{X}_t$ (i.e., $T_t = (N(N+1))^{2t} + 2$). $\mathbf{X}_t$ is a compact metric space space with the usual Euclidean distance metric, $d_t(X_t, X') = \left(\sum_{i=1}^{T_t} (X_t'(i) - X_t(i))^2\right)^{1/2}$

Let us now construct the product space of the $\mathbf{X}_t$'s

$$\mathbf{X} = \prod_{t=1}^{\infty} \mathbf{X}_t$$

Clearly the sequence $\{c_t(\theta^t), l_t(\theta^t), x_t, K_{t+1}\}_{t=0}^{\infty}$ must belong to $\mathbf{X}$. In fact, it must belong to the subset of $\mathbf{X}$, which satisfy (10), (11), and (12), denoted by $\bar{\mathbf{X}}$.

Now by Tychonoff's theorem (e.g., Dudley, 2002, Theorem 2.2.8), $\mathbf{X}$ is compact in the product topology. Since (10), (11), and (12) are (weak) inequalities, $\bar{\mathbf{X}}$ is a closed subset of $\mathbf{X}$, and therefore it is also compact in the product topology. Moreover, $\mathbf{X}$ with the product topology is meterizable, with the metric

$$d(X, X') = \sum_{t=1}^{\infty} \phi^t d_t(X_t, X_t') \tag{80}$$

for some $\omega \in (0,1)$ and $X \equiv \{X_t\}_{t=0}^{\infty} \in \mathbf{X}$. This shows that $\mathbf{X}$ endowed with the product topology is a metric space, and so is $\bar{\mathbf{X}}$. From Theorem 8, the set of probability measures defined over a compact metric space is compact in the weak topology. This establishes that the set of probability measures $\mathcal{P}^{\infty}$ defined over $\bar{\mathbf{X}}$ is compact in the weak topology.

We are concerned not with all probability measures, but those that condition at t on information revealed up to t. Let $\mathcal{C} = \{(c,l) \in \mathbb{R}^2 : 0 \leq c \leq \bar{c},\ 0 \leq l \leq \bar{l}\}$ be the set of possible consumption-labor allocations for agents, so that $\mathcal{P}^{\infty}$ defined above is the set of all probability measures over $\mathcal{C}^{\infty}$. Now, for each $t \in \mathbb{N}$ and $\theta^{t-1} \in \Theta^{t-1}$, let $\mathcal{P}\left[\theta^{t-1}\right]$ be the space of $N+1$-tuples of probability measures on Borel subsets of $\mathcal{C}$ for an individual with history of reports θ^{t-1}. Thus each element $\zeta\left(\cdot \mid \theta^{t-1}\right) = [\zeta(\theta_0 \mid \theta^{t-1}), ..., \zeta(\theta_N \mid \theta^{t-1})]$ in a $\mathcal{P}^t\left[\theta^{t-1}\right]$ consists of $N+1$ probabilities measures for each type θ_i given their past reports, θ^{t-1}, and is thus closed. Consider $\mathcal{P} \equiv \bigcup_{t \in \mathbb{N}} \bigcup_{\theta^t \in \Theta^t} \mathcal{P}^t\left[\theta^{t-1}\right]$, which is a closed subset of $\mathcal{P}^{\infty}$. Since a closed subset of a compact space is compact (e.g., Dudley, 2002, Theorem 2.2.2), $\mathcal{P}$ is compact in the weak topology.

Finally, choosing $\phi \leq \beta$ in (80) shows that the objective function is continuous in the weak topology. This establishes that including randomizations, we have a maximization problem over probability measures in which the objective function is continuous in the weak topology, and the constraint set is compact in the weak topology, and thus there exists a probability measure that reaches the maximum.

Given this result, the rest of the analysis parallels that of Theorem 3. The key lemma is a generalization of Lemma 4, which is stated here.

Lemma 11 *$\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is continuous and concave on Λ^{∞}, nondecreasing in C_s and nonincreasing in L_s for any s and differentiable in $\{C_t, L_t\}_{t=0}^{\infty}$.*

This lemma can be proved along the lines of Lemma 4, except that in this infinite-dimensional space we are no longer able to prove Lemma 3. Thus, all the proofs assume that the solution is at a regular point as defined in Definition 6.

Proof. The above argument established that in the problem of maximizing (9) subject to (10), (11), and (12) over probability measures, a maximum exists and $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is therefore well defined.

To show concavity, consider (C^0, L^0) and (C^1, L^1) and corresponding ζ^0, ζ^1. We have

$$\begin{aligned}
& \int (u(c,l;\theta) - u(c,l;\hat{\theta}))\zeta^{\alpha}(d(c,l),\theta) \\
= \ & \alpha \int (u(c,l;\theta) - u(c,l;\hat{\theta}))\zeta^0(d(c,l),\theta) + (1-\alpha)\int (u(c,l;\theta) - u(c,l;\hat{\theta}))\zeta^1(d(c,l),\theta) \\
\geq \ & 0
\end{aligned}$$

In a similar way we can show that ζ^{α} satisfies (10), (11), and (12), this convex combination is feasible and it gives the same utility as $\alpha\zeta^0 \cdot u(\theta) + (1-\alpha)\zeta^1 \cdot u(\theta)$.

Next, note that the constraint set expands if C_s increases or L_s decreases for any s, therefore U must be weakly increasing in C_s and weakly decreasing in L_s.

Finally, Theorem 7 applies to this problem and implies that $U(\{C_t, L_t\}_{t=0}^{\infty})$ is differentiable in $\{C_t, L_t\}_{t=0}^{\infty}$, completing the proof. ∎

Lemma 12 *Λ^{∞} is compact and convex.*

Proof. **Convexity:** Consider $\{C_t, L_t\}_{t=0}^{\infty}$ and $\{C'_t, L'_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$ and some ζ^0, ζ^1 feasible for $\{C_t, L_t\}_{t=0}^{\infty}$ and $\{C'_t, L'_t\}_{t=0}^{\infty}$ respectively. Now for any $\alpha \in (0,1)$ $\zeta^{\alpha} \equiv \alpha\zeta^0 + (1-\alpha)\zeta^1$ is feasible for $(\alpha \{C_t, L_t\}_{t=0}^{\infty} + (1-\alpha)\{C'_t, L'_t\}_{t=0}^{\infty})$, so that this set is non-empty. Moreover, since ζ^0, ζ^1 satisfy (56), ζ^{α} satisfies it as well. Similarly, ζ^{α} satisfies (57) and (58).

Compactness: For any sequence $\{C_t^n, L_t^n\}_{t=0}^{\infty} \in \Lambda^{\infty}, \{C_t^n, L_t^n\}_{t=0}^{\infty} \to \{C_t^{\infty}, L_t^{\infty}\}_{t=0}^{\infty}$, there exists a sequence $\{\zeta^n\}_{t=0}^{\infty}$ corresponding to $\{C_t^n, L_t^n\}_{t=0}^{\infty}$, such that $\zeta^n \to \zeta^{\infty}$, satisfying (56)-(58) and feasibility, therefore $\{C_t^{\infty}, L_t^{\infty}\}_{t=0}^{\infty} \in \Lambda^{\infty}$ is closed. Boundedness follows from boundedness of C and L. ■

13.2 Proof of Theorem 4

The proof of Theorem 4 is similar to that of Theorem 3, except that we do not use the recursive formulation. Instead, we work directly with the sequence problem and the necessary conditions of the sequence problem.

Proof. **Part 1:** Suppose to obtain a contradiction that $\mu_t(h^t) = 0$ for all $t \geq 0$ and all h^t (almost surely). Then, $x_t(h^t) = 0$ for all t and all h^t. But in this case, if $L_t > 0$ for any t, then the government can improve by expropriating the entire output at t. Thus we must have $L_t(h^t) = 0$ for all t and all h^t. Since, by hypothesis, $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \text{Int}\Lambda^{\infty}$ with $L_t(h^t) > 0$ is feasible and the associated $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \text{Int}\Lambda^{\infty}$ necessarily gives higher ex ante utility to citizens than $L_t(h^t) = C_t(h^t) = 0$, the plan with $L_t(h^t) = 0$ for all t and all h^t cannot be optimal, establishing a contradiction and proving Part 1 of the theorem.

Part 2: Take $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t)\}_{t=0}^{\infty}$ to be part of the optimal mechanism. Suppose that $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t)\}_{t=0}^{\infty}$ (almost surely) converges to a limit, (C^*, L^*, K^*), and let $x^* = L^* - C^* - K^*$.

We start by proving that $\varphi = \sup\{\varrho \in [0,1] : \text{plim}_{t\to\infty} \varrho^{-t} U_{C_t}^* = 0\}$ defined in the theorem is well-defined and strictly less than 1. To see this, recall that by hypothesis, a steady state exists, so that $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t)\}_{t=0}^{\infty} \to (C^*, L^*, K^*)$, thus $\{C_t(h^t)\}_{t=0}^{\infty}$ is in the space c of convergent infinite sequences (rather than simply in the space of all bounded infinite sequences, ℓ_{∞}). The dual of c is ℓ_1, that is, the space of sequences $\{y_t\}_{t=0}^{\infty}$ such that $\sum_{t=0}^{\infty} |y_t| < \infty$. Since $\mathcal{U}_{C_t}$ is equal to the Lagrange multiplier for the constraint (15), it lies in the dual space of $\{C_t(h^t)\}_{t=0}^{\infty}$ (see, e.g., Luenberger, 1969, Chapter 9), thus in ℓ_1, which implies that $\lim_{t\to\infty} \mathcal{U}_{C_t} = 0$, hence $\varphi < 1$.

The rest of the proof of Part 2 will consist of two cases.

Case 1: Suppose that $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \to (C^*, L^*) \in \text{Int}\bar{\Lambda}^{\infty}$ and that $\varphi = \delta$. Then (40) and (41) are necessary conditions for optimality. Rearranging these equations and substituting for $\mathcal{U}_{C_t}$, we have

$$-\frac{\mathcal{U}_{L_t}^*}{\mathcal{U}_{C_t}^* F_{L_t}(K^*, L^*)} = 1 - \frac{(\mu_t - \mu_{t-1})v'(F(K^*, L^*)))}{\mu_t v'(x^*)} \tag{81}$$

and

$$\frac{F_{K_{t+1}}(K^*, L^*)\mathcal{U}_{C_{t+1}}^*}{\mathcal{U}_{C_t}^*} = 1 + \frac{(\mu_{t+1} - \mu_t)v'(F(K^*, L^*))F_{K_{t+1}}(K^*, L^*)}{\mu_t v'(x^*)}, \tag{82}$$

where all derivatives are evaluated at the limit (C^*, L^*, K^*).

Since $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \to (C^*, L^*) \in \text{Int}\bar{\Lambda}^{\infty}$, equation (44) also holds as $t \to \infty$ and implies that, we have

$$\frac{\mathcal{U}_{C_t}^*}{\delta^t v'(x^*)} = \mu_t \leq \mu_{t+1} = \frac{\mathcal{U}_{C_{t+1}}^*}{\delta^{t+1} v'(x^*)}. \tag{83}$$

Since as $t \to \infty$, a steady state (C^*, L^*, K^*, x^*) exists by hypothesis and $\mathcal{U}_{C_t}^*$ is proportional to φ^t, (83) can be written as

$$\frac{\varphi^t \mathcal{U}_{C^*}^*}{\delta^t v'(x^*)} = \mu_t \leq \mu_{t+1} = \frac{\varphi^{t+1}\mathcal{U}_{C^*}^*}{\delta^{t+1} v'(x^*)} \text{ as } t \to \infty. \tag{84}$$

Since $\varphi = \delta$ and since, as established above, $\varphi < 1$, we have that (84) implies that as $t \to \infty$, $|\mu_{t+1} - \mu_t| \to 0$ and $\mu_t \to \mu^* \in (0, \infty)$ (where the fact that $\mu^* > 0$ follows from Part 1, since $\mu_{t+1} \geq \mu_t$

and $\mu_t > 0$ for some t). Therefore, $(\mu_t - \mu_{t-1})/\mu_t \to 0$, and from (81) and (82), we have that $-\mathcal{U}_{L_t}/\mathcal{U}_{C_t}F_{L_t}$ and $F_{K_{t+1}}\mathcal{U}_{C_{t+1}}/\mathcal{U}_{C_t}$ almost surely converge to 1, thus $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t)\}_{t=0}^{\infty}$ must be asymptotically undistorted. The rest of the argument parallels the proof of Part 2 of Theorem 3. This establishes the result for *Case 1* where $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \to (C^*, L^*) \in \text{Int}\bar{\Lambda}^{\infty}$ *and* $\varphi = \delta$.

Case 2: Suppose that either $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \to (C^*, L^*) \in \text{Bd}\bar{\Lambda}^{\infty}$ (and $\varphi = \delta$) or $\varphi < \delta$.

First consider the subcase where $\varphi < \delta$. Then, equation (83) cannot apply. Since this equation must hold for all $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \in \text{Int}\bar{\Lambda}^{\infty}$, we must have that $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \to (C^*, L^*) \in \text{Bd}\bar{\Lambda}^{\infty}$. This implies that we converge to some $\{C_t(h^t), L_t(h^t)\}_{t=0}^{\infty} \to (C^*, L^*) \in \text{Bd}\bar{\Lambda}^{\infty}$ and $\{C_t(h^t), L_t(h^t), K_{t+1}(h^t)\}_{t=0}^{\infty} \to (C^*, L^*, K^*)$ (since by hypothesis a steady state exists), as in the first part of the hypothesis. Thus we only have to consider the case where $(C^*, L^*) \in \text{Bd}\bar{\Lambda}^{\infty}$ with and $\varphi \le \delta$.

Now let us turn to the subcase where $(C^*, L^*) \in \text{Bd}\bar{\Lambda}^{\infty}$ (and $\varphi \le \delta$). Suppose also that $v(x^*)/(1-\delta) < \bar{w}$, where $\bar{w}$ is as defined in Assumption 7. Now (40) and (41) no longer apply, but since $v(x^*)/(1-\delta) < \bar{w}$, the first-order necessary conditions with respect to x_t and K_{t+1} imply:

$$\delta\mu_t v'(x_t) - \delta^t \kappa_t = 0$$

$$-\delta^t(\mu_{t+1} - \mu_t)v'(F(K_{t+1}, L_{t+1}))F_{K_{t+1}} + \delta^{t+1}\kappa_{t+1}F_{K_{t+1}} - \delta^t\kappa_t = 0,$$

where recall that κ_t is the multiplier on the resource constraint (39). Rearranging these equations by eliminating κ_t, we obtain:

$$\frac{\mu_{t+1}}{\mu_t} = \frac{\frac{v'(x_t)}{\delta F_{K_{t+1}}} - v'(F(K_{t+1}, L_{t+1}))}{v'(x_{t+1}) - v'(F(K_{t+1}, L_{t+1}))} = \frac{\frac{v'(x^*)}{\delta F_{K_{t+1}}(K^*, L^*)} - v'(F(K^*, L^*))}{v'(x^*) - v'(F(K^*, L^*))}, \tag{85}$$

where the second equality evaluates the expression at the limit (C^*, L^*, K^*, x^*). Now to obtain a contradiction, suppose that as $t \to \infty$, we have $\mu_t \to \infty$, and thus $\mu_{t+1}/\mu_t > 1$. Inspection of (85), combined with the fact that $v'(x_{t+1}) - v'(F(K_{t+1}, L_{t+1})) \ge 0$ (since $F(K_{t+1}, L_{t+1}) \ge x_{t+1}$ and $v(\cdot)$ is concave), implies that this is only possible if $\delta F_{K_{t+1}}(K^*, L^*) < 1$. However, we have $\mathcal{U}_{C_t} \le \mathcal{U}_{C_{t+1}} \cdot F_{K_{t+1}}$, thus as $t \to \infty$, $\varphi F_{K_{t+1}} \ge \delta F_{K_{t+1}} \ge 1$, where the inequality follows from the assumption that $\varphi \ge \delta$ and contradicts $\delta F_{K_{t+1}}(K^*, L^*) < 1$. Consequently, we have $\mu_t \to \mu^* < \infty$ at the limit point (C^*, L^*, K^*, x^*), so that (12) is slack, and therefore the solution to problem (19)-(20) coincides with the solution to problem (17)-(18) with $X(h^t) \to x^*$.

Finally suppose that $(C^*, L^*) \in \text{Bd}\bar{\Lambda}^{\infty}$ but also $v(x^*)/(1-\delta) = \bar{w}$. Assumption 7 then implies that (12) is slack, and once again the solution to problem (19)-(20) coincides with the solution to problem (17)-(18) with $X(h^t) \to x^*$. This completes the proof of *Case 2* and thus of Part 2.

Part 3: Suppose that $\varphi > \delta$. In this case, (83) implies that $\mathcal{U}^*_{C_t}$ is proportional to $\varphi > \delta$ as $t \to \infty$. This implies that $(\mu_t - \mu_{t-1})/\mu_t > 0$ as $t \to \infty$, so from (81) and (82), aggregate distortions cannot disappear, completing the proof. ■

14 Appendix E: Proofs for Section 7

14.1 Proof of Theorem 5:

The proof of this theorem follows the structure of the proofs of Lemma 1, Theorem 1 and Proposition 1.

Proof. As in the proof of Lemma 1, we first need to show that there exists a sequentially rational continuation play in which all agents supply zero labor. Suppose that the government has announced a submechanism $\tilde{M}_t$ at time t and has capital stock K_t, and $\alpha^i_{t+s} = \alpha^{\varnothing}$ for all $i \in [0,1]$ and for all $s \ge 0$. We first show that a deviation by an individual, i' with type $\theta^{i'}_t \ne \theta_0$ to some other strategy that involves supplying positive labor is not profitable (as noted in footnote 10, we think of an individual with positive measure ε deviating, and take the limit $\varepsilon \to 0$, since there is a continuum of agents). Without the deviation, i' obtains utility $u(0)/(1-\beta)$ (since from Assumption 1', $\chi(0 \mid \theta) = 0$ for all $\theta \in \Theta$

and there will be no labor supply for any type in the continuation game). Now imagine a deviation to a message that corresponds to positive labor supply, say l', with $\chi\left(l' \mid \theta_t^{i'}\right) > \chi\left(0 \mid \theta_t^{i'}\right) = 0$ by definition. This will generate output $F(K_t, \varepsilon l')$, since all other agents are supplying zero labor. Now imagine the behavior of the government at the last stage of the game, conditional on $\alpha_{t+s}^i = \alpha^{\varnothing}$ for all $i \in [0,1]$ and for all $s \geq 1$. Then the sequentially rational strategy of the government is to maximize (49) with $K_{t+1} = 0$, since there will be no production in future periods. Consequently, the utility-maximizing program of the government in the information set following the deviation is (suppressing h^t-dependence to simplify notation):

$$\max_{\tilde{x}_t', \tilde{c}_t'} (1-a) v(x_t) + a \left(\int \left[u\left(\tilde{c}_t'\left(z^t\left(\alpha_t\left(\theta^t\right)\right)\right)\right) - \chi\left(l_t\left(z^t\left(\alpha_t\left(\theta^t\right)\right)\right) \mid \theta_t\right) \right] dG^t\left(\theta^t\right) \right),$$

subject to $x_t + \int \tilde{c}_t'\left(z^t\left(\alpha_t\left(\theta^t\right)\right)\right) dG^t\left(\theta^t\right) \leq F(K_t, \varepsilon l')$, where recall that $z^t\left(\alpha_t\left(\theta^t\right)\right)$ is the history of reports up to time t by an individual of type θ^t given strategy profile $\underline{\alpha}$. In view of Assumption 1', this expression is concave in c for any strategy profile $\underline{\alpha}$, so the optimal policy for the government in this information set is to redistribute consumption (what it does not consume itself) equally across agents, i.e., $\tilde{c}_t'\left(z^t\left(\alpha_t\left(\theta^t\right)\right)\right) = c_t$ for all $z^t\left(\alpha_t\left(\theta^t\right)\right) \in Z^t$. This implies that as $\varepsilon \to 0$, $c_t \to 0$, and thus the deviation payoff of i' is $u(0) - \chi\left(l' \mid \theta^{i'}\right) + \beta\left(u(0) - \chi\left(0 \mid \theta^{i'}\right)\right)/(1-\beta) < \left(u(0) - \chi\left(0 \mid \theta^{i'}\right)\right)/(1-\beta)$, showing that a continuation strategy profile where all agents supply zero labor is sequentially rational.

Now consider two different types of deviations by the government. First, imagine the government offers $\tilde{M}_t \neq M_t$, i.e., a different mechanism at the beginning of time t than the one implicitly agreed in the social plan (M, x). Given the above-constructed continuation equilibrium, $\alpha_{t+s}^i = \alpha^{\varnothing}$ for all $i \in [0,1]$ and for all $s \geq 0$ is a best response against this deviation. Since maximal punishments are optimal, $\alpha_{t+s}^i = \alpha^{\varnothing}$ for all $i \in [0,1]$ and for all $s \geq 0$ is optimal against this deviation, implying that such a deviation would never be profitable for the government.

Second, as before, the government can deviate at the last stage of time t. Again $\alpha_{t+s}^i = \alpha^{\varnothing}$ for all $i \in [0,1]$ and for all $s \geq 1$ is the maximal sequentially rational punishment against such a deviation. Consequently, after any deviation by the government, there will not be any further production. Thus the optimal deviation for the government involves $\tilde{K}_{t+1}' = 0$, and again exploiting the concavity of the government's continuation payoff in c, the sustainability constraint is equivalent to:

$$\begin{aligned} & \mathbb{E}_t \sum_{s=0}^{\infty} \delta^s \left[(1-a) v(x_{t+s}) + a \mathbb{E}_{t+s} \left(\int \left[u(c_t) - \chi\left(l_t\left(z^t\left(\alpha_t\left(\theta^t\right)\right)\right) \mid \theta_t\right) \right] dG^t\left(\theta^t\right) \right) \right] \quad (86) \\ \geq & \max_{\tilde{x}_t' + \tilde{c}_t' \leq F(K_t, L_t)} (1-a) v(\tilde{x}_t) + a \int u\left(\tilde{c}_t'\left(\theta^t\right) dG^t\left(\theta^t\right)\right) \text{ for all } t. \end{aligned}$$

Now, given an equilibrium pair of strategy profiles Γ and $\underline{\alpha}$, exactly the same argument as in the proof of Theorem 1 implies that there exists another pair of equilibrium strategy profiles Γ^* and $\underline{\alpha}^* = (\alpha^* \mid \alpha')$ for some α' such that Γ^* induces direct submechanisms. Consequently, we can write (86), in terms of a direct mechanism, which gives (50).

Finally, the same argument as in the proof of Proposition 1 implies that the best sustainable mechanism is a solution to maximizing (9) subject to (10), (11), and the sustainability constraints of the government given by (50). ■

14.2 Proof of Theorem 6

Proof. Suppose again that there are $N+1$ types, i.e., $\Theta = \{\theta_0, \theta_1, ..., \theta_N\}$, ranked in ascending order of skills, and with respective probabilities $\{\pi_0, \pi_1, ..., \pi_N\}$. Given the assumptions of the theorem (and again suppressing h^t-dependence to simplify notation), we can write the program for the best sustainable

mechanism as:

$$\max_{\left\{\{c_t(\theta_i),l_t(\theta_i)\}_{i=0}^{N},x_t,K_{t+1}\right\}_{t=0}^{\infty}} \sum_{t=0}^{\infty}\beta^t\sum_{i=0}^{N}\pi_i\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]$$

subject to the constraints

$$\sum_{t=0}^{\infty}\beta^t\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]\geq\sum_{t=0}^{\infty}\beta^t\left[u\left(c_t\left(\theta_{i-1}\right)\right)-\chi\left(l_t\left(\theta_{i-1}\right)\mid\theta_i\right)\right] \tag{87}$$

for all $i=1,\ldots,N$,

$$\sum_{s=0}^{\infty}\beta^{t+s}\left\{(1-a)\,v\left(x_{t+s}\right)+a\left(\sum_{i=0}^{N}\pi_i\left[u\left(c_{t+s}\left(\theta_i\right)\right)-\chi\left(l_{t+s}\left(\theta_i\right)\mid\theta_i\right)\right]\right)\right\}\geq V\left(K_t,L_t\right) \tag{88}$$

for all t, and

$$x_t+K_{t+1}+\sum_{i=0}^{N}\pi_i u\left(c_t\left(\theta_i\right)\right)\leq F\left(K_t,\sum_{i=1}^{N}\pi_i l_t\left(\theta_i\mid\theta_i\right)\right) \tag{89}$$

for all t, and that $c_t\left(\theta_i\right)\geq 0$ for all i and t and $x_t\geq 0$ for all t.

The first set of constraints, (87), ensure incentive compatibility for the citizens. Given Theorem 5, there is truthful revelation along the equilibrium path. This, together with Assumption 2, implies that we only need one constraint for each type other than the disabled type, θ_0, where type i could deviate to claim to be type $i-1$. The second set of constraints, (88), one for each date, impose sustainability, with the definition $V\left(K_t,L_t\right)\equiv\max_{\tilde{x}_t'+\tilde{c}_t'\leq F(K_t,L_t)}(1-a)\,v\left(\tilde{x}_t'\right)+a\sum_{i=0}^{N}\pi_i u\left(\tilde{c}_t'\left(\theta_i\right)\right)$, and finally, the last set of constraints, one for each date, impose the aggregate resource constraint.

As in subsection 6.1, we follow Marcet and Marimon (1998) and form the Lagrangian (see footnote 41 for details):

$$\begin{aligned}
\mathcal{L} \;=\; & \sum_{t=0}^{\infty}\beta^t\sum_{i=1}^{N}\pi_i\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right] \\
& +\sum_{t=0}^{\infty}\beta^t\mu_t\left\{(1-a)\,v(x_t)+a\sum_{i=1}^{N}\pi_i\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]\right\} \\
& -\sum_{t=0}^{\infty}\beta^t(\mu_t-\mu_{t-1})V\left(K_t,\sum_{i=1}^{N}\pi_i l_t\left(\theta_i\mid\theta_i\right)\right) \\
& +\sum_{i=1}^{N}\lambda_i\left\{\sum_{t=0}^{\infty}\beta^t\left\{\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]-\left[u\left(c_t\left(\theta_{i-1}\right)\right)-\chi\left(l_t\left(\theta_{i-1}\right)\mid\theta_i\right)\right]\right\}\right\} \\
& -\sum_{t=0}^{\infty}\beta^t\eta_t\left\{x_t+K_{t+1}+\sum_{i=1}^{N}\pi_i c_t\left(\theta_i\right)-F\left(K_t,\sum_{i=1}^{N}\pi_i l_t\left(\theta_i\mid\theta_i\right)\right)\right\}
\end{aligned}$$

where λ_i is the multiplier on the incentive-compatibility constraint of type i, $\beta^t\eta_t$ is the multiplier on the resource constraint at time t, and we have left the constraints that $c_t\left(\theta_i\right)\geq 0$ for all i and t and $x_t\geq 0$ for all t implicit.

For $c_t\left(\theta_i\right)>0$ and $x_t>0$, we can take first-order conditions, which, after canceling out the β^t terms and defining $\lambda_0=0$ and $\lambda_{N+1}=0$, yield:

$$\left(1+a\mu_t\right)\pi_i u'\left(c_t\left(\theta_i\right)\right)+\left(\lambda_i-\lambda_{i+1}\right)u'\left(c_t\left(\theta_i\right)\right)-\eta_t\pi_i=0 \text{ for all } i=0,\ldots N \text{ and all } t, \tag{90}$$

$$
\begin{aligned}
& (1+a\mu_t)\,\pi_i\chi'\left(l_t(\theta_i)\mid\theta_i\right)+\left(\lambda_i\chi'\left(l_t(\theta_i)\mid\theta_i\right)-\lambda_{i+1}\chi'\left(l_t(\theta_i)\mid\theta_{i+1}\right)\right) \\
+ \quad & \left(\mu_t-\mu_{t-1}\right)\pi_i V_L\left(K_t,L_t\right)-\eta_t\pi_i F_L\left(K_t,L_t\right)=0 \text{ for all } i=0,\dots N \text{ and all } t,
\end{aligned} \tag{91}
$$

$$
-\left(\mu_t-\mu_{t-1}\right)V_K\left(K_t,L_t\right)+\eta_t F_K\left(K_t,L_t\right)-\beta^{-1}\eta_{t-1}=0 \text{ for all } t, \tag{92}
$$

$$
\mu_t\left(1-a\right)v'\left(x_t\right)=\eta_t \text{ for all } t. \tag{93}
$$

Recall that $\{\mu_t\}_{t=0}^{\infty}$ is a nondecreasing sequence, thus it possesses a unique limit point on the extended real line. First suppose that $\mu_t \to \mu^* < \infty$, then (90)-(93) establishes that there exists an allocation with $\eta_t \to \eta^*$, $c_t(\theta_i) \to c_i^* > 0$ for all i, and $x_t \to x^*$, in which distortions disappear as claimed in the theorem.

To complete the proof, we need to show that there does not exist any solution to the above maximization problem with $\mu_t \to \infty$. To obtain a contradiction suppose that $\mu_t \to \infty$. Combine (90) for type N with (93) and using the fact that $\mu_{t+1} \geq \mu_t > 0$ and that $\lambda_{N+1} = 0$ (by definition), we have that for all $c_t(\theta_N) > 0$ and $x_t > 0$,

$$
\frac{\left(1+\frac{\lambda_N}{\pi_N}\right)}{(1-a)\frac{v'(x_t)}{u'(c_t(\theta_N))}-a}=\mu_t\leq\mu_{t+1}=\frac{\left(1+\frac{\lambda_N}{\pi_N}\right)}{(1-a)\frac{v'(x_{t+1})}{u'(c_{t+1}(\theta_N))}-a}. \tag{94}
$$

Both sides of this equation are strictly positive by the fact that $\mu_{t+1} \geq \mu_t > 0$. The hypothesis that $\mu_t \to \infty$ implies that as $t \to \infty$, $\mu_t < \mu_{t+1}$.

Next combine (90) some i and $i' \neq i$ to obtain:

$$
u'\left(c_t(\theta_i)\right)+\frac{(\lambda_i-\lambda_{i+1})}{\pi_i\left(1+a\mu_t\right)}u'\left(c_t(\theta_i)\right)=u'\left(c_t(\theta_{i'})\right)+\frac{(\lambda_{i'}-\lambda_{i'+1})}{\pi_{i'}\left(1+a\mu_t\right)}u'\left(c_t(\theta_{i'})\right). \tag{95}
$$

The fact that $\mu_t \to \infty$ implies that as $t \to \infty$, $|c_t(\theta_i) - c_t(\theta_{i'})| \to 0$. This argument then establishes that $c_t(\theta_i) \downarrow c^*$ for all $i = 0, \dots N$. From the freedom of labor supply, this also implies that we must have $l_t(\theta_i) \downarrow l^* = 0$ for all $i = 1, \dots N$, since otherwise at some point all $\theta_i \neq \theta_0$ would claim to have become disabled). From Assumption 2 and the resource constraint, this also implies that as $t \to \infty$, $c_t(\theta_i) \downarrow 0$ for all $i = 0, \dots N$ and $x_t \downarrow 0$.

Now, suppose first that $l^* = 0$ and $c^* = 0$ are reached in finite time, i.e., $c_{t'}(\theta_i) = 0$ for all i and all $t \geq t'$ for some $t' < \infty$. We will show that this cannot be part of a sustainable mechanism. We have that for all $t \geq t'$, $c_t(\theta_i) = l_t(\theta_i) = 0$ for all i and $x_t = 0$, so the continuation utility of the government is $(1-a)\,u(0)/(1-\delta)$ (since $v(0) = 0$ and $\chi(0 \mid \cdot) = 0$). However, by hypothesis, $c_{t'-1}(\theta_i) > 0$ for at least some i, so there is positive output at $t'-1$. Moreover, from (95), $c_{t'-1}(\theta_i) \neq c_{t'-1}(\theta_{i'})$ for some i and i'. This implies that the government would prefer to deviate at $t'-1$ to $\xi_{t'-1} = 1$, and redistribute the output between x_t and equal consumption across all individuals (i.e., $c_{t'-1}(\theta_i) = c^*_{t'-1}$ for all i and some $c^*_{t'-1}$). This deviation will necessarily increase government utility at $t'-1$ (since $c_{t'-1}(\theta_i) \neq c_{t'-1}(\theta_{i'})$ for all i and i' in the original allocation), and its continuation utility from t' onwards would still remain at $(1-a)\,u(0)/(1-\delta)$. Since this argument applies for any $t' > 0$, it proves that there cannot be a sustainable mechanism that reaches $l^* = 0$ and $c^* = 0$ in finite time. Hence, it must be the case that $c_t(\theta_i) \downarrow 0$ and $l_t(\theta_i) \downarrow 0$, but $c_t(\theta_i) > 0$ for all t. Then, combining (94) and (95) implies that, as long as $au'(0) \neq (1-a)\,v'(0)$, $\mu_{t+1} - \mu_t \downarrow 0$, contradicting $\mu_t \to \infty$, and thus establishing the theorem. ■

15 References

Abreu, Dilip (1988) "On the Theory of Repeated Games with Discounting" *Econometrica*, 56, 383-396.

Acemoglu, Daron (2005a) "Modeling Inefficient Institutions" forthcoming in *Proceedings of 2005 World Congress of the Econometric Society*, edited by R. Blundell, W. Newey and T. Persson.

Acemoglu, Daron (2005b) "Politics and Economics in Weak and Strong States" *Journal of Monetary Economics*, 52, 1199-1226.

Acemoglu, Daron, Michael Golosov and Aleh Tsyvinski (2006) "Tax Policies of the Leviathan" in progress.

Acemoglu, Daron and Simon Johnson (2005) "Unbundling Institutions" *Journal of Political Economy*, 113, 949-995.

Acemoglu, Daron, Michael Kremer and Atif Mian (2003) "Incentives in Markets, Firms and Governments" NBER working paper No. 9802.

Albanesi, Stefania and Christopher Sleet (2005) "Dynamic Optimal Taxation with Private Information", *Review of Economic Studies*, 72, 1-29

Atkeson, Andrew and Robert E. Lucas, Jr. (1992) "On efficient distribution with private information", *Review of Economic Studies,* 59, 427-53.

Ausubel Lawrence and Raymond Deneckere (1989) "Reputation in Bargaining and Durable Goods Monopoly" *Econometrica*, 57, 511-531.

Austen-Smith, David, and Jeffrey S. Banks (1999) *Positive Political Theory I: Collective Preference.* Ann Arbor MI: University of Michigan Press.

Baron, David and Roger Myerson (1982) "Regulating a Monopolist with Unknown Costs " *Econometrica*, 50, 911-30.

Battaglini, Marco and Stephen Coate (2005) "Pareto Efficient Taxation with Stochastic Abilities" Cornell and Princeton mimeo.

Benveniste, Lawrence and José Scheinkman (1979) "On the Differentiability of the Value Function in Dynamic Models of Economics" *Econometrica*, 47, 727-732.

Bertsekas, Dimitri, Angelia Nedic and Asuman Ozdaglar (2003) *Convex Analysis and Optimization*, Athena Scientific Boston.

Bester, Helmut and Richard Strausz (2001) "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case." *Econometrica*, 69, 1077-1098.

Bonnans, J. Frederic and Alexander Shapiro (2000) *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer, New York.

Bisin, Alberto and Adriano Rampini (2005) "Markets as Beneficial Constraints on the Government", forthcoming *Journal of Public Economics*

Buchanan, James M. and Gordon Tullock (1962) *The Calculus of Consent.* Ann Arbor, MI: University of Michigan Press.

Chamley, Christophe (1986) "Optimal Taxation of Capital Income in General Equilibrium and Infinite Lives" *Econometrica*, 54, 607-622.

Chari, V.V. (2000) "Limits of Markets and Limits of Governments: An Introduction to a Symposium on Political Economy" *Journal of Economic Theory*, 94, 1-6.

Chari, V.V. and Patrick Kehoe (1990) "Sustainable Plans" *Journal of Political Economy*, 94, 783-802.

Chari, V.V. and Patrick Kehoe (1993) "Sustainable Plans and Mutual Default" *Review of Economic Studies*, 60, 175-195.

Dasgupta, Partha, Peter Hammond and Eric Maskin (1979) "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility", *Review of Economic Studies*, 46(2),185-216

Dixit, Avinash K. (2004) *Lawlessness and Economics: Alternative Modes of Governance*, Princeton; Princeton University Press.

Dudley, Richard (2002) *Real Analysis and Probability*, New York, Cambridge University Press.

Freixas, Xavier, Roger Guesnerie and Jean Tirole (1985) "Planning under Incomplete Information and the Ratchet Effect" *Review of Economic Studies*, 52, 173-192.

Fudenberg, Drew, David Levine and Eric Maskin (1994) "The Folk Theorem with Imperfect Public Information" *Econometrica*, 62, 997-1039.

Golosov, Michael, Narayana Kocherlakota and Aleh Tsyvinski (2003) "Optimal Indirect and Capital Taxation", *Review of Economic Studies* 70, 569-587

Golosov, Michael and Aleh Tsyvinski (2004) "Designing Optimal Disability Insurance: A Case for Asset Testing", NBER Working Paper #10792

Green, Edward J. (1987) Lending and the smoothing of uninsurable income, in *Contractual Arrangements for Intertemporal Trade,* ed. E. Prescott and N. Wallace, Minneapolis: University of Minnesota Press, 3-25.

Green Jerry and Jean Jacques Laffont (1977) "Characterization of Satisfaction Mechanisms for to Revelation of Preferences for Public Goods" *Econometrica*, 45, 427-38.

Harris, Milton and Robert M. Townsend (1981) "Resource Allocation Under Asymmetric Information", *Econometrica*, 49(1), 33-64.

Hart, Oliver, Andrei Shleifer and Robert Vishny (1997) "The Proper Scope of Government: Theory and an Application to Prisons," *Quarterly Journal of Economics*, November, 112(4), 1127-1161.

Holmstrom Bengt and Roger Myerson (1983) "Efficient and Durable Decision Rules with Incomplete Information" *Econometrica*, 51, 1799-1819.

Judd, Kenneth (1985) "Redistributive Taxation in a Simple Perfect Foresight Model" *Journal of Public Economics*, 28, 59-83.

Kocherlakota, Narayana (2005) "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation", *Econometrica*, 73(5), 1587-1621.

Luenberger, David G. (1969) *Optimization by Vector Space Methods*, John Wiley & Sons New York.

Marcet, Albert and Ramon Marimon (1998) "Recursive Contracts". Mimeo. University of Pompeu Fabra

Mas-Colell, Andreu, Michael D. Whinston and Jerry R. Green (1995) *Microeconomic Theory*, Oxford University Press, New York, Oxford

Miller, David (2005) "Attainable Payoffs in Repeated Games with Interdependent Private Information" University of California San Diego mimeo.

Mirrlees, James A. (1971) "An Exploration in the Theory of Optimal Income Taxation", *Review of Economic Studies* 38, 175-208.

Myerson, Roger B. (1979) "Incentive Compatibility and the Bargaining Problem", *Econometrica*, 47(1), 61-73

North, Douglass C. and Robert P. Thomas (1973) *The Rise of the Western World: A New Economic History*, Cambridge University Press, Cambridge UK.

North, Douglass C. and Barry R. Weingast (1989) "Constitutions and Commitment: Evolution of Institutions Governing Public Choice in Seventeenth Century England, *Journal of Economic History*, 49, 803-832.

Olson, Mancur (1982) *The Rise and Decline of Nations: Economic Growth, Stagflation, and Economic Rigidities,* Yale University Press, New Haven and London.

Parthasarathy, K. R. (1967) *Probability Measures on Metric Spaces*, Academic Press, New York.

Persson, Torsten and Guido Tabellini (2000) *Political Economics: Explaining Economic Policy*, The MIT Press, Cambridge MA

Phelan, Christopher (1994) "Incentives and Aggregate Shocks" *Review of Economic Studies*, 61, 681-700.

Phelan, Christopher and Ennio Stacchetti (2001) "Sequential Equilibria in a Ramsey Tax Model", *Econometrica*, 69(6), 1491-1518.

Phelan, Christopher and Robert Townsend (1991) "Computing Multi-Period, Information-Constrained Optima" *Review of Economic Studies*, 58, 583-881.

Prescott, Edward C. and Robert Townsend (1984a) "General Competitive Analysis in an Economy with Private Information", *International Economic Review*, 25, 1-20

Prescott, Edward C. and Robert Townsend (1984b) "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard", *Econometrica* 52, 21-45.

Ray, Debraj (2002) "Time Structure of Self-Enforcing Agreements" *Econometrica*,70, 547-82.

Roberts, Kevin (1977) "Voting Over Income Tax Schedules" *Journal of Public Economics*, 8,329-40.

Roberts, Kevin (1984) "Theoretical Limits to Redistribution" *Review of Economic Studies*, 51, 177-195.

Rockafellar, R. Tyrrell (1970) *Convex Analysis*, Princeton University Press, Princeton.

Segal, Ilya (2004) "The Communication Requirements of Social Choice Rules and Supporting Budget Sets" Stanford University mimeo.

Skreta, Vasiliki (2004) "Sequentially Optimal Mechanisms," *NAJ Economics,* Vol. 4 - March 05, 2002, http://www.najecon.org/v4.html.

Sleet, Christopher and Sevin Yeltekin (2004) "Credible Social Insurance" Carnegie Mellon mimeo.

Stokey, Nancy, Robert E. Lucas and Edward Prescott (1989) *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge.

Thomas, Jonathan and Tim Worrall (1990) "Income Fluctuations and Asymmetric Information: An Example of Repeated Principle-Agent Problem", *Journal of Economic Theory* 51, 367-90.

Uhlig, Harald (1996) "A Law of Large Numbers for Large Economies" *Economic Theory*, 8, 41-50.

Werning, Ivan (2002) "Optimal Dynamic Taxation and Social Insurance", University of Chicago Ph.D. dissertation.